

Don't believe the hype

Nanotechnologists are increasingly concerned about the lurid descriptions of the dangers of their work being promulgated by environmental campaigners. But the field's proponents aren't helping their cause by making exaggerated claims.

Anyone who has given a public lecture on nanotechnology has probably been asked the question: "What about the 'grey goo'?" The term, coined by the futurist Eric Drexler back in the 1980s, describes the ever-growing mess of self-replicating nanorobots as they consume the environment and manufacture more of their own kind. For researchers working on the science of the incredibly small, this highly improbable scenario has become a tiresome piece of apocalyptic baggage.

The tiresome may be about to become genuinely threatening, however. Seasoned environmental campaigners, fresh from bloodying the nose of the biotechnology industry over the issue of genetically modified (GM) crops, are now eyeing nanotechnology as their next target (see page 246). Some are even calling for a moratorium on research in the field. And they've cooked up a new goo with which to scare the public: the marriage of nanotechnology and biotechnology to create new forms of life that will behave in unpredictable and uncontrollable ways, threatening natural habitats, biodiversity and human health — 'green goo', as it has been dubbed.

Researchers weary from years of grey-goo scaremongering may throw up their hands at the prospect of a second, similarly improbable form of gunk. Certainly, the discussion at a meeting held in June in Brussels, organized by the Action Group on Erosion Technology and Concentration (ETC) of Winnipeg, Canada, to stoke political opposition to nanotechnology in Europe, veered into the downright ludicrous. One activist argued that potatoes could be constructed from nanorobots, threatening traditional methods of agriculture.

Faced with such patent nonsense, it is easy to dismiss the anti-nanotech campaign as an irrelevant rant — easy, but very unwise. The claims made by groups such as ETC may at times appear irrational, but that doesn't mean they won't have resonance with the public. Many campaigners against biotech and nanotech aren't concerned so much about the specific dangers of the two technologies as about the general environmental and social consequences of advanced technologies when placed in the hands of multinational corporations. Many ordinary citizens, and some scientists, have similar misgivings about this trend towards economic techno-globalization. As a result, groups such as ETC may gain a sympathetic ear.

Public concern

What's more, the European experience with GM crops has shown that environmental campaigners are adept at building from vague public misgivings to whip up a storm of protest, even if their arguments rest on shaky scientific ground. In Britain, for instance, where 'mad cow disease' had fed distrust of official assurances about food safety, green campaigners in 1999 skilfully exploited a study that purported to show that rats fed GM potatoes suffered health problems. The research was deeply flawed, but the misleading message that GM food can kill touched a public nerve in a way that the most thorough investigation into the effects of GM agriculture on farmland biodiversity — an issue of genuine scientific controversy — never could.

Many of the activists now calling for strict controls on nanotechnology are veterans of the GM crop wars. ETC, for instance,

was formerly known as the Rural Advancement Foundation International, which enjoyed considerable success in its efforts to portray Monsanto and other agribiotech companies as enemies of the developing world's small-scale farmers. In making an explicit link with agribiotech by deploying its green-goo scenario, ETC may succeed in transferring public concerns about GM crops to nanotechnology, even though the issues involved are different.

How should those working on nanoscale science respond? Simply dismissing the claims of ETC and its allies out of hand, and arguing that nanotechnology poses no conceivable hazards, will do little to win hearts and minds. Many people remember similar assurances from the past: that mad cow disease posed no threat to human health, and that transgenes would not find their way from GM crops into wild relatives, for example.

Honest debate

Another tactic, already being used by some researchers, is to argue that there's nothing really new about nanotechnology: it's simply an extension of existing research in chemistry, materials science, physics and engineering. This is also likely to fail because nanotechnologists — or, at least, their official spokespeople — have spent the past few years telling the public just the opposite. New results from the field of nanoscale science are paraded into the media spotlight almost daily, each allegedly capable of curing disease, cleansing the environment, or otherwise extending peace and prosperity. Such hype ensures that nobody will believe the message that nanotech is old news. And if the technology is capable of such wonders, people may muse, the chances are that it can probably do bad things as well.

So what can researchers do to address public concerns about nanotechnology? First, don't panic. The European backlash against GM crops provides a gloomy backdrop, but the conditions for the storm were created by a spectacular own goal by the US biotechnology industry. The decision not to segregate exports of soya beans into GM and non-GM batches enraged European public opinion by denying consumer choice. Such mistakes don't have to be repeated.

Second, scientists should engage the public in honest debate about the potential risks posed by nanotechnology, and how they can be managed. One of the rallying calls of the anti-nanotech campaign is that too little is being done to address the toxicity of engineered nanoparticles. According to preliminary studies, nanoparticles can disperse through the environment and may do some harm inside the body. Scientists should start thinking about how to limit people's exposure to them.

More research into the risks posed by nanoparticles is warranted, and the findings must be shared with the public. At the same time, researchers should engage in debate about the understandable fear that lies behind green goo — that new technologies can radically alter our unstable world. In doing so, they need to counter the hype currently being deployed in generous measure by both opponents and proponents of nanotechnology. The challenge is to convey the possibilities and risks of this new science without painting it as a panacea or a plague. ■





Up in flames?
Physicists warn of
problems with US
missile-defence plan
p240



HIV negative
United States diverts
research money from
AIDS to anthrax
p241



Top of the world
Arctic rocket team
sets sights on the
atmosphere's heights
p243



Test result
‘Trick’ nerve-gas
studies will not
face legal action
p244

China launches primate centre to broaden medical use of monkeys

David Cyranoski, Tokyo

Biologists in China plan to capitalize on the region's easy access to primates to set up a global resource for disease research.

The centre, based at Sun Yat-sen University in Guangzhou, southern China, will use locally obtained primates to build up a library of stem cells. It will also develop large colonies of transgenic primates for use as models of human disease.

Together, Sun Yat-sen University and local government have invested about US\$1 million in the project, which is tentatively named the Center for Stem Cell Biology and Tissue Engineering.

The centre is being established in close cooperation with Bruce Lahn, a geneticist at the University of Chicago, Illinois. It aims to explore areas in which the use of primates has so far been limited, such as neurological disease and developmental biology. In the United States and Europe, primate research in these areas is restricted because of the high cost and lengthy life cycle of the animals, as well as criticism from animal-rights activists.

Cost should not be a problem for the new centre, says its director Peng Xiang, a former postdoctoral student of Lahn's. US researchers often pay US\$5,000 per animal, but the centre will pay about one-tenth of that, Xiang says, obtaining several hundred rhesus macaques and crab-eating monkeys from a primate breeding centre nearby.



Rhesus monkeys are among the primates that Chinese researchers will use to model human disease.

Xiang is also looking at obtaining inbred animals from an island south of China, although this may prove to be controversial as the removal of monkeys from the wild has upset conservationists (see *Nature* **417**, 684–687; 2002).

Lahn believes that the centre could re-define primate research. “China's unique advantages in the field could propel it into a position of international leadership,” he says.

If so, it would bring many experiments, currently done in mice, one step closer to humans. “Primate research has tremendous promise for bridging the gap between what's known about the biology of mice and what we want to know about human disease,” says Gerald Schatten of the University of Pittsburgh, Pennsylvania, who has pioneered transgenic and cloning studies in primates.

Xiang argues that many experiments

Are China's bioethics under control?

China has similar protocols on biological experimentation to other countries, researchers there often argue, but it is less bureaucratic in implementing them.

As recent health problems with AIDS and SARS have demonstrated, however, the Chinese government may lack the means to enforce strict regulations throughout the vast country, even if it wanted to.

As a result, research ethics are being policed at the local level — or they rely on the goodwill of China's researchers, and their appetite for international acceptance.

“Our first priority is to publish papers in the

major journals, and we don't want to get rejected on the basis of a violation of ethical standards,” says Linsong Li, director of Peking University's Stem Cell Research Center.

The Chinese government has been preparing ethical guidelines for biologists for more than a year, but their publication has repeatedly been delayed. The document is expected to cover human stem cells and to include some restrictions on other procedures, such as xenotransplantation.

In the meantime, some local organizations have taken matters into their own hands. The National Human Genome Center in Shanghai, for example, in spring 2002 drafted its own guidelines

for research on human embryonic stem cells.

Even if major centres have regulations in place, it is hard to know what is happening in the rest of China. Although human rights are taken seriously, the idea of humane treatment of animals has yet to be cemented, says Huanming Yang, director of the Beijing Genomics Institute. “We're still at the beginning with animal rights,” he adds.

Chinese biologists are highly sensitive to the outside perception that their country has lax bioethical standards, says Bruce Lahn of the University of Chicago. Many Chinese research leaders were trained abroad and follow international rules, he argues.

David Cyranoski

using mouse models have been inconclusive because of the great dissimilarity between humans and mice.

Already, 15 researchers are working at the new centre. According to Lahn, researchers will start with 100–200 monkeys this autumn and build to three times that number over the next few years. The founders of the centre say that they will adhere to Western-level research ethics, although doubts have sometimes been expressed about China's ability to police such ethics (see 'Are China's bioethics under control?', previous page).

The centre is expected to build on some of Schatten's pioneering work in primates, including efforts to clone a rhesus monkey (C. Simerly *et al. Science* **300**, 297; 2003), and a gene-transfer experiment that achieved low-level expression of a jellyfish gene for a green fluorescent protein in another rhesus monkey (A. W. S. Chan, K. Y. Chong, C. Martinovich, C. Simerly and G. Schatten *Science* **291**, 309–312; 2001).

According to Xiang, the centre will try to improve on previous transgenic experiments by using several different viruses to introduce the genes. The viruses will either be allowed to invade a newly fertilized egg on their own or will be injected directly into an egg, as the researchers look for the best way to create primates with active transgenes. The resulting animals could be used as models for neurodegenerative diseases such as Huntington's, Xiang says. The centre will also study the function of developmental genes such as *hedgehog*, focusing in particular on their role in brain development.

Another of the centre's aims is to establish itself as a global source of primate stem cells of various types and stages of development. The supply of primate stem cells will "be valuable in a few years, when clinical trials of stem-cell therapies are closer to reality and real preclinical work is necessary", says Daniel Salomon of the Scripps Research Institute in La Jolla, California.

But Salomon doubts whether the centre's plan to create a pool of inbred monkeys will provide a useful model for disease research. "Inbred mice have been extraordinarily useful for basic studies," he says, "but the translation to human patients that are everything but inbred has been problematic."

The new centre's managers are already wary of ethical charges that may be levelled against their work. They have backed away, for example, from an earlier plan to grow tissue from human cells in primates for implantation back into humans. Such 'xenotransplantation' is seen by many experts as dangerous because of the risks of interspecies infection. ■



Blast off: plans to use short-range interceptor missiles for US defence face technical challenges.

Experts find fault with US plan to intercept missiles at source

Geoff Brumfiel, Washington

Defending the United States against long-range missile attack may be even more technically challenging than was previously thought, a group of physicists has warned.

In a study due to be published this week, a team from the American Physical Society examined the possibility that enemy missiles could be destroyed within four minutes of launching. The idea has support in US military and scientific circles, as missiles are most visible as targets during this period.

The physicists conclude that interceptor missiles will need to be placed close to enemy territory to reach a missile in this early phase. Even then, the study says, decision-makers will have only seconds to decide whether to launch an interceptor. And if the missile destroys its target, the warhead might survive and fall into a civilian area. "The problems are extremely difficult," says Daniel Kleppner, a physicist at the Massachusetts Institute of Technology (MIT), and co-chair of the study.

The United States spends about \$8 billion a year on developing systems for shooting down incoming missiles. Most of this goes to interceptors being installed at Fort Greely, Alaska, which are designed to collide with a warhead at the apex of its flight. But many previous studies have cast doubt on this strategy, as the warheads are small and difficult to hit, and the interceptors can be fooled by balloons and other decoys. Many experts have argued that it would be better to stop missiles at the start of their flight, when the warhead is attached to fuel tanks, creating a larger target.

The physicists show, however, that interceptor missiles would have to be placed

within 400–1,000 km of the launch site to hit their target. For small, coastal countries such as North Korea, interceptors could be fired from US ships. But action against countries such as Iran would require neighbouring states to host US missile bases. Interceptors would also need to reach speeds of up to 10 km per second — much faster than current systems.

A more vexing problem is the risk that warheads from intercepted missiles could hit civilian targets. "A nuclear warhead is designed to be very robust," says Ted Postol, an MIT physicist and missile-defence expert. Although the chances of hitting a populated area are small, Postol argues that the warhead is likely to land in a country not involved in any conflict.

The study also looks at two unconventional proposals for early interception of ballistic missiles: an airborne laser that would fire at enemy missiles as they take off, and space-based interceptors that would descend on a missile. The laser could be effective if it were sited within 600 km of the launch site, say the authors, but the space system would be impossible to implement without about 1,000 interceptors in orbit. That would require at least a "five- to tenfold" increase in the capacity of US space-launch systems.

The report may not change many minds on this partisan topic, but it will inform the debate over whether to catch missiles as they take off, says Philip Coyle, senior adviser to the Center for Defense Information in Washington DC. "This study is the first honest-to-God assessment of what's scientifically possible," he says. ■

♦ www.aps.org

AIDS research cut to pay for anthrax vaccine

Erika Check, Washington

The Bush administration is to proceed with plans to skim more than \$200 million from research grant programmes to pay for the rapid production of an anthrax vaccine, brushing off the protests of biologists.

As a result of the decision, 375 AIDS researchers and other grant-holders at the National Institute of Allergy and Infectious Diseases (NIAID) will this year lose the initial six months of funding on their awards.

The NIAID will spend \$233 million on the research, development and purchase of a 'next-generation' anthrax vaccine by 2004, the White House Office of Management and Budget says.

The transfer of money from the NIAID's civilian research programmes for the vaccine work is just 2% of its total research budget. But research organizations say they are afraid that the decision is a harbinger of how the NIAID's swelling biodefence mission may compromise its research programmes.

"I don't think anyone opposes doing research to make a better anthrax vaccine — but that work should be funded out of the bioterrorism budget, not by raiding the AIDS budget," says Daniel Kuritzkes, director of AIDS research at the Partners AIDS Research Center in Cambridge, Massachusetts.

Congress and the administration have been wrangling over the anthrax-vaccine



At the sharp end: anthrax vaccines for US troops are a high priority for the Bush administration.

project since February last year, when President Bush's budget request asked for \$233 million for the NIAID to spend on a vaccine. Congress denied the request and divided up the money between several parts of the National Institutes of Health. But the White House then demanded that the NIAID find a way to fulfil its request anyway.

In a letter sent to legislators on 2 July, White House budget director Joshua Bolten said that the NIAID would spend up to \$117 million this year and \$116 million next year on the "advanced development" of an anthrax vaccine, including the purchase of up to 9 million doses of vaccine.

The decision disappointed the Infectious Diseases Society of America, which says that the vaccine purchase could endanger the NIAID's larger research mission. The group argues that another branch of government, such as the Department of Homeland Security, should be paying for the vaccine.

On 11 July, Congressman Henry Waxman (Democrat, California) and Senator Jeff Bingaman (Democrat, New Mexico) wrote to the president protesting against the decision, which they called a "serious mistake".

But NIAID officials are putting on a brave face. "The Office of Management and Budget's position is that there is a critical need for the nation to rapidly develop a vaccine and there's nothing else out there to support this now," says Ralph Tate, the NIAID's budget director.

Janet Shoemaker, public-affairs director at the American Society for Microbiology, says NIAID officials are making the best of a difficult situation. "The anthrax issue has become less urgent in most people's minds, but in the minds of the people making the decisions it is still a very high priority," she says. ■

Time runs short for opponents of chemical-testing rules

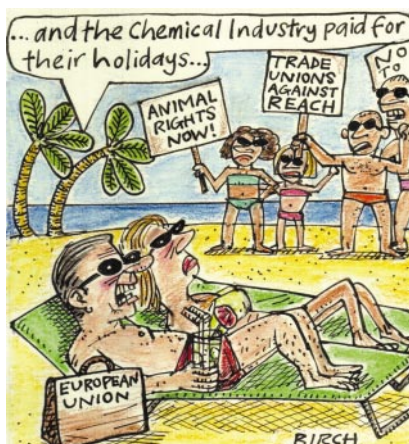
Glenn Murphy, London

An unlikely alliance of animal-rights organizations and chemical-industry bodies is running out of time in its bid to stop a proposed chemical-testing plan from being implemented in Europe.

An eight-week consultation on the latest draft of the European Commission's new policy, which campaigners warn could lead to a massive increase in testing on animals, ended last week. Opponents of the legislation say they doubt whether any major changes will be made before the proposals go before the European Parliament later this year.

The draft contains details for a new Registration, Evaluation and Authorisation of Chemicals (REACH) system for all untested chemicals manufactured in the European Union (EU) in quantities over one tonne. The system would reverse the burden of proof, requiring manufacturers to demonstrate the safety of a new product.

The rules are being introduced in part to guide the assessment of existing chemicals, as European-level testing only began in 1981. Campaigners say that the process



would involve the evaluation of more than 30,000 chemicals and could potentially take decades. Animal-rights groups estimate that more than 12 million animals would be needed (see *Nature* 418, 116; 2002).

Representatives from the chemical industry, trade unions and animal-rights organizations, who met in London last week to publicize their concerns, complain that

the rules make no attempt to prioritize chemicals by looking at probable risks and existing data. The result, they say, will be duplication of tests and the redundant testing of everything from benzene to baking soda. Similar arguments have surrounded the introduction of a new testing regime in the United States (see *Nature* 395, 828; 1998).

The campaigners warn that chemical firms may relocate outside Europe.

Chemicals imported to the EU will need to have REACH clearance, but products containing chemicals may not.

Commission officials say that a new testing system is badly needed in the interest of public health and that the legislation can still be amended. But campaigners fear that the draft will be rushed through before the European Parliament's term ends next May.

This doesn't bode well for those who want to change the rules, says Roger Jeary, national secretary of manufacturing for the London-based trade union Amicus. "Once you get to the stage of publishing draft legislation," he says, "the changes tend to be at the edges, rather than in the substance." ■

Seed bank raises hope of Iraqi crop comeback

Tom Clarke, London

Agricultural scientists in Syria are sitting on a 'black box' of seed samples that could help plant research to get back on its feet in Iraq — if the security situation there improves.

The collection of more than 200 seed samples, removed from Iraq in 1996 and stored at the International Center for Agricultural Research in the Dry Areas (ICARDA) in Aleppo, should ease the restoration of agricultural research in Iraq, researchers say.

But a seed bank will be worthless without a functioning agricultural sector to use it. At a press conference in Washington DC on 10 July, Dan Amstutz, the former agribusiness executive appointed by the US Department of Agriculture to reconstruct Iraqi farming, admitted that the deteriorating security situation had forced that effort into retreat. "We viewed getting out and talking to farmers as an enormously important part of our function," he said, "and that's more difficult now."

In the short term, restoring Iraq's agriculture hinges on ensuring that farmers have enough seed to plant and the means to harvest it. But the genetic diversity held in seed banks is crucial to long-term farm productivity, experts say. Seeds that confer resistance to pests, diseases or drought but are also adapted to the local environment are essential for sustainable production.

None of the 1,400 crop varieties stored in Iraq's Abu Gharib national gene bank, near Baghdad, is thought to have survived the aftermath of this year's invasion by the United States and Britain, says William Erskine, ICARDA's research director.

But seven years ago, scientists at Abu Gharib, alarmed by deteriorating conditions in their own country, sent samples to ICARDA



Despite the ravages of sanctions and war, hopes remain that Iraq's agricultural prowess can be revived.

for safe-keeping. "Someone at Abu Gharib showed great foresight," says Erskine. The collection contains more than 200 of the most valuable varieties of 28 different crops.

Plant scientists hope that other varieties can also be saved. Iraq occupies an area that was probably the first in the world to refine the agriculture of such crops as wheat and barley — making local strains particularly interesting to plant scientists globally. This makes it more likely that the remaining 1,200 varieties in Abu Gharib's original collection can eventually be located. "Seeds from the region were well collected in the past," says Mike Ambrose, a crop geneticist at the John Innes Centre in Norwich, UK. "I'd hope that a lot of them can be repatriated," he says.

Researchers are also keen to salvage Iraq's own crop-science prowess, which remained considerable despite its erosion during the time of United Nations sanctions. "We need to get back quickly to prevent a diaspora of scientific expertise," says Erskine.

In May, the US Agency for International Development began inviting applications from US companies or non-profit organizations to begin rebuilding Iraq's agricultural infrastructure, and for partnerships between US and Iraqi universities to bolster research. No contract has yet been awarded — the largest restoration job to date has been an \$800,000 US project to restore basic amenities at the agriculture ministry's looted headquarters in Baghdad.

Ants join online colony to boost conservation efforts

Rex Dalton, San Francisco

An Internet-based system of ant images is expected to help ecologists study and preserve global biodiversity.

The website, www.antweb.org, provides researchers with three-dimensional images of thousands of ant species, along with all of the information held in traditional catalogues of species.

"This sets new standards for doing taxonomy online," says Brian Fisher, an entomologist at the California Academy of Sciences in San Francisco, who led the \$30,000 project to set up the system.

AntWeb already holds images of many of the world's 10,000 species, including those from California (270 species) and Madagascar (210 species), where more than



Small wonder: *Microdaceton tibialis* is one of hundreds of species in the AntWeb catalogue.

1,000 species are under threat from deforestation. Six hundred species, mainly from Florida, will be added to the website by

the end of the year. The eventual goal is to include all of the world's ants.

Ants are important in many ecosystems, and AntWeb is intended to help conservation biologists by providing quick access to information on every ant they come across.

The Nature Conservancy, for example, plans to use AntWeb to help its ecologists manage the Lassen Foothills, an ecologically sensitive region of about 900,000 hectares in California. "The evaluation of our conservation efforts is a great challenge," says Mark Reynolds, a Nature Conservancy ecologist based in San Francisco. "AntWeb gives us the ability to know what is happening at the invertebrate level."

Arctic rockets give glimpse of the atmosphere's top layers

M. RAPP

Quirin Schiermeier, Spitsbergen

A series of rocket-borne atmospheric reconnaissance missions in the Arctic Circle is yielding potentially valuable data on the little-known thermodynamics of the middle and upper atmosphere.

On 8 July a German–Norwegian team launched the third and final mission in the ROMA (Rocket-borne Observation in the Middle Atmosphere) project. The probe blasted off from an international research base at Ny Ålesund, on the Norwegian island of Spitsbergen.

ROMA's rocket-borne, returnable probes measure temperature, density and ion concentration on their ascent to altitudes of about 105 kilometres. The €250,000 (US\$280,000) instruments are launched on 6.5-metre, solid-fuelled rockets — which researchers inherited when they were discarded by the Norwegian military — and splash down in the Barents Sea, after which they are retrieved for analysis.

The middle and upper atmosphere is poorly understood, compared with the well-investigated troposphere and stratosphere. “We know little about turbulence, or about fluctuations in density and temperature,” says ROMA's project leader Tom Blix, a space physicist at the Norwegian Defence Research Establishment in Kjeller, near Oslo.

The middle atmosphere is the coolest region of the Earth's atmosphere, reaching –160 °C in polar regions in a layer called the mesopause, at an altitude of around 90 kilometres. Above this layer, temperatures rise again, and density decreases, through the ionosphere.

All of these layers are coupled, however, and the current global warming in the lower layers is expected to resonate throughout the atmosphere. A better understanding of the chemical and physical processes that take place at high altitudes may therefore help researchers to model long-term variations in climate.

But it is difficult to study the middle and upper atmosphere from the ground. Some researchers have sought to do so using pulsed laser instruments called LIDARs (light detection and ranging). The reflections of these lasers provide some information — about ice-particle concentrations, for example — but do not tell researchers much about thermodynamic and electromagnetic processes that occur high in the sky.

“Only rocket-borne *in situ* measurements provide the highly resolved vertical snapshots that allow us to look for atmospheric fluctuations on all scales,” says Blix.

Ny Ålesund is the northernmost perma-



Sky high: the final rocket in the ROMA project prepares for blast-off from its Arctic launchpad.

nently inhabited village in the world. It is an ideal location for probing the upper atmosphere, says Markus Rapp, a physicist at the Leibniz Institute of Atmospheric Physics in Kühlungsborn, Germany, and a member of the ROMA team.

At the village's latitude of 79 °N, the force lines of the Earth's magnetic field are almost perpendicular to the Earth's surface, allowing ions from the solar wind to enter the atmosphere.

The ions have long been known to cause the northern lights, or *aurora borealis*, which are commonly observed at northern latitudes in winter. Free electrons, density fluctuations and the occurrence of ice clouds in the upper atmosphere also combine to produce radar phenomena called polar mesosphere summer echoes. These effects have only recently been explained (M. Rapp and F.-J. Lübken *J. Geophys. Res.* doi:10.1029/2002JD002857; 2003).

The project is also tracking ionized ice particles in the upper atmosphere, which form ‘noctilucent clouds’ that are visible at dawn and dusk in modestly high latitudes, where the ice clouds scatter sunlight from below the horizon. Researchers believe that noctilucent clouds are linked to atmospheric radar echoes observed at high latitudes in summer.

ROMA's preliminary data have surprised its participants. “Density fluctuations and turbulence at high altitudes appear to be much stronger than we had expected,” says Rapp.

But the cause of these fluctuations, and their impact on atmospheric science, have yet to emerge — the ROMA team is only now beginning to analyse its results. ■

Preliminary inquiry clears brain scientist of fraud allegation

Alison Abbott, Munich

An official investigation has cleared a prominent German neuroscientist of manipulating data in one of two scientific papers under scrutiny.

The allegations against Heinz Breer, a molecular neuroscientist at the University of Hohenheim in Stuttgart, were made earlier this year, prompting an investigation by the DFG, Germany's main funding agency. Breer says that his reputation has already been damaged by reports in German newspapers about the allegations (see *Nature* 422, 794; 2003).

The DFG's investigating committee this month issued an interim report on one of the two papers co-authored by Breer (S. Schreiber, J. Fleischer, H. Breer and I. Boekhoff *J. Biol. Chem.* 275, 24115–24123; 2000). It ruled that the paper contains “deficiencies of a technical nature”, although it attributes these to sloppiness rather than an intention to mislead.

The paper is a study of the molecular processes by which the brain interprets odours. One apparent error identified by the committee relates to the analysis of data on the accumulation of a signalling molecule under different experimental conditions. The second involved mislabelling of negative controls in western blots — a method for identifying proteins separated by chromatography.

Breer, who won the prestigious Leibniz research award in 1998, has previously conceded minor problems with the paper. The experiments were mostly carried out by Ingrid Boekhoff, a young researcher in Breer's lab, and by an undergraduate student. The committee's report says the case “makes clear the importance of making young scientists familiar with the principles of good scientific practice”. Boekhoff has agreed to send a correction to the *Journal of Biological Chemistry*.

Wieland Huttner, a molecular neurobiologist at the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden, is one of two reviewers appointed by the DFG to investigate the case. He says that some of the data preparation was “sloppy”, but that there is no evidence of cheating. He adds that the publicity surrounding the case, including German newspaper reports of the allegations, has been “unfair” to both Breer and Boekhoff.

The DFG committee is expected to rule on the second paper (J. Noé and H. Breer *J. Neurochem.* 71, 2286–2293; 1998) in the next few months. ■

NASA on board as search for intelligent life takes off again

Washington The hunt for distant civilizations is back on NASA's agenda. The SETI (Search for Extraterrestrial Intelligence) Institute, based in Mountain View, California, announced earlier this month that it had received a five-year, US\$5-million grant from NASA to conduct research into the planetary context for life.

The money represents SETI's first major funding from NASA since 1993. Funds were cut after critics in Congress caricatured SETI as a wacky hunt for aliens. The institute continued as a private non-profit organization, but the recent economic downturn has put it under increasing financial pressure.

"We're glad to have them on board," says Rose Grymes, who heads NASA's Astrobiology Institute at the Ames Research Centre at Moffett Field, California, which is funding the research. Grymes says that SETI researchers will conduct studies of Jupiter's moons, as well as preparing a list of stars that might be home to intelligent life.

Japanese push for patents 'sidelines basic research'

Tokyo Japanese researchers have been given their strongest instructions yet to push the industrial applications of their work.

Guidelines on the creation and protection of intellectual property, released on 8 July by a committee of politicians and external experts, call on funding agencies and ministries to give patents and intellectual property equal importance with publications when evaluating researchers.

"The new system will influence not only funding decisions for individuals but also university and research-institute evaluations, and eventually hiring decisions," says Koichi Sumikura, an intellectual-property expert at the National Graduate Institute for Policy Studies in Tokyo.

The guidelines have heightened fears that the government is more concerned with intellectual property than basic research. Only two of the committee's 28 members were scientists. Creation of intellectual property has also been a key ingredient of plans that were approved last week to reorganize the national universities (see *Nature* 419, 875–876; 2002).

Nerve-gas experiments will not go to court

London No prosecutions will be made over controversial human research that took place between 1939 and 1989 at Porton Down, a UK Ministry of Defence (MOD)



In safe hands? Volunteers allege that they were unfairly experimented on at Porton Down.

weapons testing facility in Wiltshire.

The decision not to bring charges, announced by the Crown Prosecution Service on 8 July, follows a four-year inquiry carried out by Wiltshire police and the MOD into allegations of illegal experiments carried out at the facility.

Operation Antler involved interviews with 250 ex-service personnel, many of whom claim they were tricked into participating in nerve-gas experiments — some under the pretence of research into the common cold — and had suffered long-term illness as a result.

Eight cases were passed to the prosecution service by investigators, but the service said that more than 700 people were contacted in the investigation, and although many had suffered illness, others subjected to the same experiments showed no long-term effects.

A separate coroner's inquest into the death of Ronald George Maddison, who died during testing at Porton Down in 1953, has been set for September.

Grants plan paves way for Europe-wide funding body

Munich The European Molecular Biology Organization (EMBO) is to become a grant-giving agency.

The change provides a stimulus to those who want to create a Europe-wide science funding body. "This could act as a prototype for other, more ambitious programmes, such as the European Research Council," says Julio Celis, a cell biologist at the University of Aarhus in Denmark and president of the European Molecular Biology Conference, the intergovernmental body that funds EMBO.

Members of the conference earlier this month agreed in principle to launch the EMBO Research Award Programme. The 24 member states will now negotiate how much each will contribute towards the €30 million (US\$34 million) that will be awarded each year. EMBO hopes to be able to issue its first call for proposals next year.

The move could also aid plans to create a European Research Area covered by a unified science policy.

Britain takes fresh look at spallation source

London The dwindling hopes of European researchers who want to host a next-generation neutron source received a boost last week, when UK science minister David Sainsbury commissioned a review of Britain's neutron-science projects.

The review will include an assessment of the proposed £1-billion (US\$1.6-billion) European Spallation Source, which would be the most powerful in the world. Plans for the source, which would probe the structures of materials and molecules, have been given a lukewarm reception by funding bodies in France and Germany (see *Nature* 418, 262; 2002). The British review, which will be completed by 2005, has revived hopes that the source could yet get built.

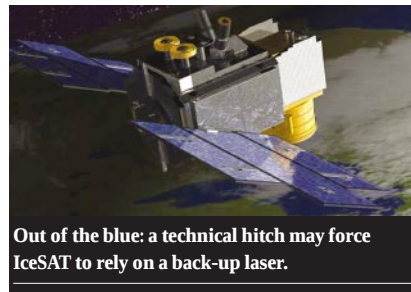
"It's tremendous news," says Robert Cywinski, a physicist at the University of Leeds and an adviser on the project. "We have the largest and most experienced community of neutron scatterers in Europe, and we need to provide facilities for their research."

Laser breakdown takes satellite's eye off the ice

Washington Data from a NASA satellite designed to monitor the mass of polar ice sheets stopped flowing this month when its laser altimeter broke down. The laser on the Ice, Cloud, and Land Elevation Satellite (ICESat), launched in January, scans ice sheets to reveal how they change over time.

Project manager Jim Watzin, based at NASA's Goddard Space Flight Center in Greenbelt, Maryland, says that ICESat has two back-up lasers, but he doesn't plan to use them straight away. "We must be sure that the back-ups will not break down in the same way," he says. A NASA panel is investigating and hopes to have a rescue plan by the end of the month.

Another troubled satellite is faring better. Scientists working on the Solar and Heliospheric Observatory have found a way to avoid the periodic data blackouts caused by a malfunctioning antenna (see *Nature* 423, 910; 2003). Signals will be sent from another antenna that is normally used to transmit non-scientific data.



Out of the blue: a technical hitch may force ICESAT to rely on a back-up laser.

A little knowledge...

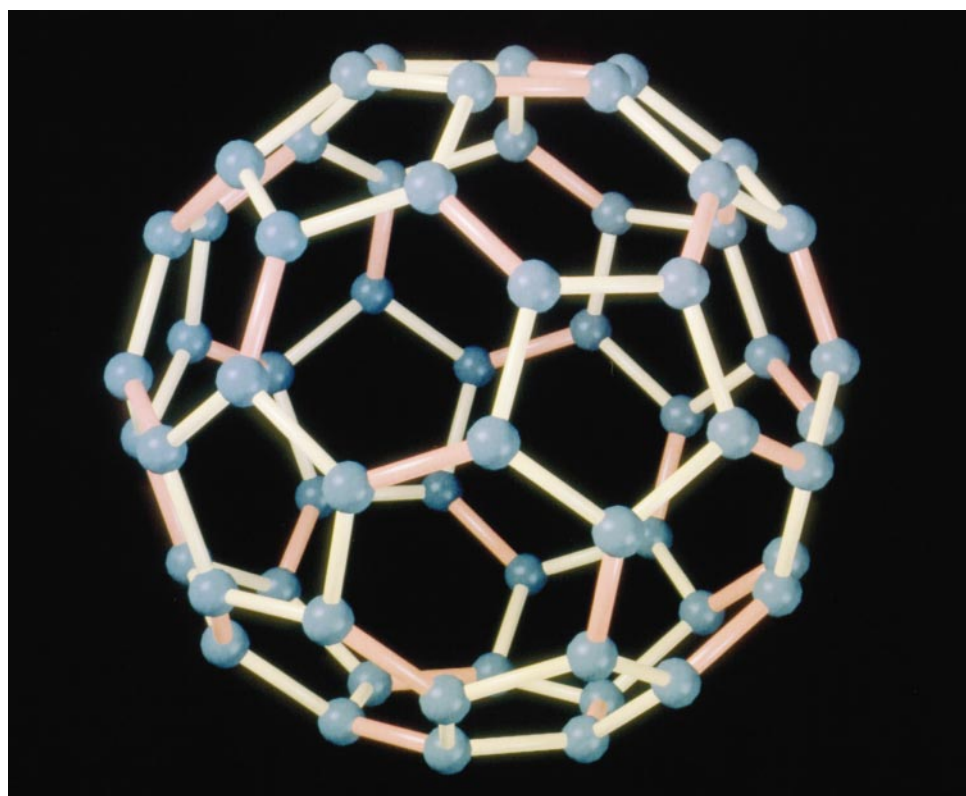
Nanotechnology is set to be the next campaign focus for environmental groups. Can scientists avoid the mistakes made over genetically modified food, and secure public trust for their research? Geoff Brumfiel investigates.

If Patrick Mooney has his way, thousands of physicists, chemists and biologists will have to down tools. His problem is nanotechnology. Right now, says Mooney, there are no rules about how nanosized particles should be dealt with in the laboratory or the market-place — the result, he argues, could be environmental pollution or even the loss of human life.

Mooney's claims certainly sound disturbing. According to the environmental organization that he heads — the Action Group on Erosion, Technology and Concentration (ETC) based in Winnipeg, Canada — nanoparticles from industry and research labs are entering the environment, the food-chain and the human body. Because little is known about the way these particles behave, says ETC, the consequences are unpredictable.

In March 2002, ETC acted on these fears, calling for a worldwide ban on nanotechnology research. Despite the organization's small size — it has just eight members of staff — the call was covered in newspapers around the globe. Generating an impact is something that ETC is good at. In its former guise as the Rural Advancement Foundation International, the group ran high-profile campaigns against bioprospecting and genetically modified foods.

Inside the labs that Mooney is targeting, many scientists wonder what all the fuss is about. Engineered nanoparticles exist in such small quantities that they could do little to harm the environment, researchers argue.



Tiny fears: environmentalists are worried about the safety of nanoparticles such as carbon 'buckyballs'.

Scientists working with the particles take common-sense measures to prevent their spread. And many types of nanoparticles are already present in our environment — diesel engines, for example, emit a range of nanosized particles, such as various hydrocarbons.

But dig a little deeper into the issues surrounding ETC's call for a moratorium, and matters become less clear. Some of the organization's claims may lack credence, but there are genuine questions to be answered about the safety of nanoparticles. And for researchers, simply dismissing people's fears is a risky strategy. Agribusinesses ignored public concern over transgenic crops during the 1990s, and many consumers now reject the technology. Researchers may not believe nanotechnology is dangerous, say experts, but they must be prepared to accept the possibility — and to confront the sometimes far-fetched claims made by environmental groups.

That shrinking feeling

Part of the difficulty with assessing ETC's claims is the fact that nanotechnology is such a broad term, covering the study and application of any material with nanometre dimensions. Hype has surrounded recent breakthroughs in the generation of nanosized wires¹ or tubes², which could one day be used to build smaller and less power-hungry computer circuits, or ultrasensitive tools for studying cell biology³. But nanotech is already being put to more mundane uses. Nanoparticles of titanium dioxide, for

example, are an important ingredient in transparent sunscreens. From everyday applications to blue-sky research, nanotechnology around the world now attracts more than US\$3 billion in investment every year.

Fears about the technology date back to the 1980s, when futurist Eric Drexler outlined a world populated by the 'grey goo' of self-replicating nanorobots that had got out of control⁴. This nightmare vision is not the main concern raised by ETC — although the group has recently raised the spectre of a destructive 'green goo' of uncontrollable life-forms created through a fusion of biotech and nanotechnology. Nor is it terribly likely to happen — the chemistry involved in building synthetic, self-replicating particles is well beyond the current techniques.

But away from the sci-fi stories, there are genuine concerns about the technology — namely the way in which nanoparticles interact with the human body and the environment. The Center for Biological and Environmental Nanotechnology at Rice University in Texas is one of the few labs to study these issues. In one small-scale project, led by environmental chemist Mason Tomson, researchers investigated how tiny cages of 60 carbon atoms, commonly called buckyballs, travel through soil. The team suspended the buckyballs in water and then poured them through a soil-like material. When the balls were allowed to clump together to form particles a few micrometres big, they were absorbed into the soil like any other organic



A. FORGET/AFP

Proponents of nanotechnology hope to avoid the protests that have dogged transgenic food.

compound. But when they were dispersed, the water formed a protective sheath around each buckyball, allowing them to travel through the soil without being absorbed. "It's completely unexpected," says Tomson.

Unpublished studies by the team show that the nanoparticles could easily be absorbed by earthworms, possibly allowing them to move up the food-chain and reach humans, says Vicki Colvin, the centre's director.

Hidden danger

Much more work needs to be done before researchers can say how dangerous these particles could be. In terms of human health, for example, it is unclear what nanoparticles would do if they entered the human body. Günter Oberdörster, a professor of environmental medicine at the University of Rochester in New York, is one of a few researchers who have been studying the effects of ultrafine particles on the body for decades. According to results he presented this May at a workshop on nanotechnology in the environment, held in Arlington, Virginia, much depends on the size of the particle involved.

Micrometre-sized clumps of nanoparticles, for example, are relatively unreactive because their surface area is smaller than that of the same number of individual nanoparticles, and they are too large to enter the bloodstream when breathed in. But individual nanoparticles can pass from the lungs into the bloodstream, and are more reactive.

For example, Oberdörster has shown that rats exposed to a mist of nanometre-sized polytetrafluoroethylene, or 'Teflon', particles experienced respiratory irritation⁵.

So conventional compounds normally considered to be harmless might prove to be dangerous on a nanometre scale — but what about the new nanoparticles being created by scientists? These, too, can be harmful, according to Chiu-Wing Lam, a senior toxicologist at NASA's Johnson Space Center in Houston, Texas, whose team has studied the health effects of carbon nanotubes. The researchers found that mice inhaling micrometre-sized clumps of tangled carbon nanotubes had the same reaction as they would to ordinary dust. But when they were exposed to individual carbon nanofibres, the mice developed lesions in their lungs and intestines. "Carbon nanotubes are not innocuous," says Lam. He believes that they should be handled only in an industrial-hygiene environment.

Some nanotechnology researchers began taking precautions before such results were produced. When Richard Smalley started mass-producing carbon nanotubes at Rice University in the late 1990s, an extension of a research programme that earned him a share of the 1996 Nobel Prize in Chemistry, he says that he looked into the risks in great detail.

Smalley's team found that nanoparticles tended to spread around the lab, and onto the clothes and the skin of people working there. "We were concerned about it," he recalls. His group set up an enclosed area, within which

they could work with the nanotubes before converting them into a powder form, which is easier to contain. Other researchers have taken similar precautions in their own labs, he says.

But beyond such self-regulation, no special rules exist for nanotechnology. This absence of control, together with the holes in our knowledge about the impact of nanoparticles, has given ETC and other environmental organizations a basis on which to campaign.

Since ETC's moratorium call, Mooney says that life has been "an absolute whirlwind". The group took its message on the road, meeting with dozens of environmentalists and public-policy groups in countries around the world. Although many environmental organizations are unsure whether to get involved, groups such as Greenpeace have begun to monitor nanotechnology research.

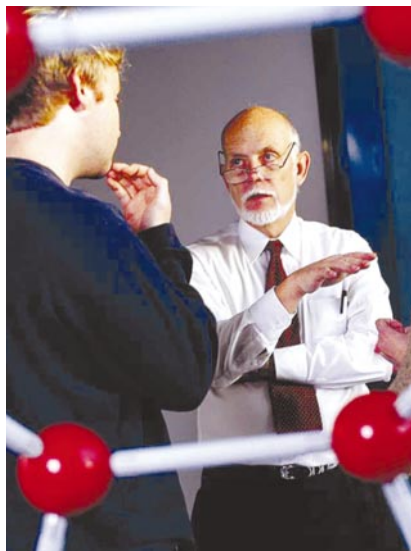
Under review

National governments have also begun to address nanotechnology. The US House of Representatives has called for more money to be spent on studying the ethical and societal implications of nanotechnology, and the Senate would like to see a centre set up to address the same issues. And last month, the British government asked the Royal Society and the Royal Academy of Engineering to investigate the potential impact of the technology. Funding is already beginning to flow into new projects — the US Environmental Protection Agency alone will spend about \$6 million this year.

Senior nanotechnology and science-policy experts have also started to consider safety issues. Industry and environmental representatives have discussed potential environmental hazards at meetings organized by David Rejeski, director of the foresight and governance programme at the Woodrow Wilson International Center for Scholars in Washington DC. In Europe, bodies such as Britain's Parliamentary Office of Science and Technology are monitoring the issue, in part because of worries that negative feelings about nanotechnology could spiral out of control, just as public concern over genetically modified crops has done (see 'Tough lessons from the fields', overleaf).

The debate is clearly gathering pace, so how should nanotechnology researchers respond? Most are confident that the benefits of nanotechnology will outweigh any harmful effects, and that their field is not that different from other areas of chemistry and physics. Nanoparticles have, for example, been used for centuries by humans in processes such as pottery glazing⁶. This leads some researchers to feel that they don't need to join in the argument. "They don't really see what the hoop-la is about," says Mark Modzelewski, executive director of the NanoBusiness Alliance, a group based in New York City that represents more than 250 US nanotech research firms.

Researchers may also find it hard to engage



For and against: nanotech pioneer Richard Smalley (right) and campaigner Patrick Mooney (left).

with what they see as unrealistic claims. In a communiqué issued this May, ETC warned that scientists eventually will be able to write DNA sequences in the same way as they do computer programs. Some nanotech advocates, tired of dispelling myths about grey goo, question whether it is worth responding to such speculation. "Why should the debate constantly be on things that most, if not all, of those in the labs think are nonsense?" asks Modzelewski. He says that there is nothing inherently wrong with ETC's scepticism, but he wants scientists to have the time to gather the proper data on safety issues.

Colvin, however, is in favour of acting now. She points out that scientists have been making nanostructures since the 1970s, but that when such research addressed environmental issues, it mainly focused on solving problems, such as how to improve solar cells. Until recently, says Colvin, researchers were reluctant to investigate the potential negative effects of nanotechnology.

She is trying to tackle this attitude by pro-

moting a wider range of work at her institute, and by encouraging researchers to engage with the public. "We're trying to teach scientists to take the argument emotionally, not technically," she says. Rather than give complex responses about why the risk is minimal, Colvin believes scientists should address the fears of the public. She tries to persuade them to talk people through what is known and admit that more research needs to be done.

Easy prey

Such public engagement may need to happen quickly. ETC's latest meeting, held in conjunction with other environmental organizations at the European Parliament in Brussels on 11 June, was attended by politicians from the Labour and Green parties in the parliament — the latter backed Mooney's moratorium call. And in last year's *Prey*⁷, best-selling author Michael Crichton once more raised the fear of grey goo, describing how biologically synthesized nanorobots wreak havoc in the United

States. The book is being made into a film. "Within weeks of its release, tens of millions will know something about nanotechnology," says Rejeski.

But will such public awareness, together with environmental campaigns, result in restrictions on nanotechnology? In terms of new legislative action, the answer is probably no. Renzo Tomellini, head of the nanosciences and nanotechnologies unit at the European Commission, says that more research on safety issues needs to be done before legislation can be considered. Newt Gingrich, former speaker of the US House of Representatives and a nanotechnology advocate, agrees. "If you look at the immediacy of SARS or smallpox, this is a lot further away," he says. "Political leaders have a limited attention span."

Nevertheless, Gingrich says that nanoscience is such an "explosive technology" that it is bound to catch the attention of government regulators at some point. When it does, researchers will need to be ready to stand up for their work. An outright ban on nanotechnology is extremely unlikely, but poorly constructed regulations could still hamper science. And of all the parties involved, say experts, researchers are in an ideal position to make sure that this doesn't happen. "Nobody has a crystal ball," says Colvin. "But scientists and engineers are best placed to guess what can go wrong."

Geoff Brumfiel is Nature's Washington physical sciences correspondent. Additional reporting by Jim Giles, Nature's associate news and features editor.

1. Appell, D. *Nature* **419**, 553–555 (2002).
2. Ball, P. *Nature* **414**, 142–144 (2001).
3. Zandonella, C. *Nature* **423**, 10–12 (2003).
4. Drexler, K. E. *Engines of Creation* (Anchor Press/Doubleday, Garden City, New York, 1986).
5. Oberdörster, G., Gelein, R. M., Ferin, J. & Weiss, B. *Inhal. Toxicol.* **7**, 111–124 (1995).
6. Padovani, S. et al. *J. Appl. Phys.* **93**, 10058–10063 (2003).
7. Crichton, M. *Prey* (HarperCollins, London, 2002).

ETC

♦ www.etcgroup.org

NanoBusiness Alliance

♦ www.nanobusiness.org

Tough lessons from the fields

If there is one thing nanotechnology advocates and most environmentalists agree on, it's that they do not want a repeat of the European debate on genetically modified (GM) foods.

Over the past 20 years, firms such as chemical company DuPont in Wilmington, Delaware, have invested millions in developing new strains of GM crops. "We went into GM foods with great enthusiasm as a business," DuPont scientist Edward Boyes told researchers at an Environmental Protection Agency meeting this spring.

But from the point of view of many European consumers, GM foods seemed to be of benefit only to farmers and agribusiness. To make matters worse for industry, companies such as Monsanto in St Louis, Missouri, were slow to respond when environmental groups began raising concerns about the safety of GM crops.

In 1997, for example, it emerged that Monsanto was refusing to separate GM and non-GM soya beans exported from the United States. Environmental groups seized on the decision, claiming

that the firm was denying consumers the right to choose not to eat GM food. A backlash developed that eventually led to several European countries blocking the licensing process for GM crops, and many supermarkets refusing to stock food made from them.

Now companies — including DuPont, which is investing heavily in nanotechnology research — are hoping to avoid these mistakes by studying the environmental risks of new technologies.

Despite these good intentions, environmentalists are sceptical. "The

lessons from biotechnology don't seem to have been learned," says Sue Mayer, director of GeneWatch, a pressure group based in Buxton, UK. GeneWatch is concerned primarily with genetic technologies, but has recently begun tracking nanotechnology. Scientists and industry representatives may be considering the negative aspects of nanotechnology, but Mayer says that the debate is insular. "If there are assumptions made about a technology, which aren't broadly shared by the public, it will cause problems," she predicts.

A new atlas of the brain

When considering the location of human cognitive functions, neuroscientists still refer to imprecise anatomical maps drawn up almost a century ago. But not for much longer, says Alison Abbott.

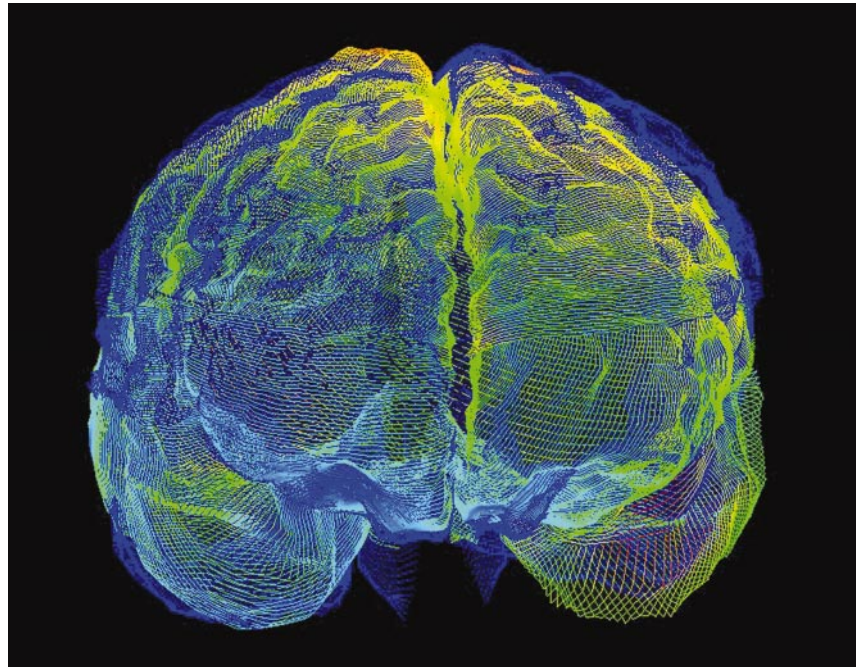
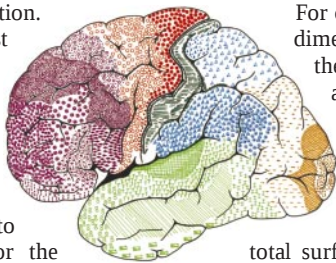
After the death of Vladimir Ilyich Lenin in 1924, the Soviet authorities wanted to locate the seat of his 'genius'. So they sought the help of Oskar Vogt of the University of Berlin, a pioneering neuroanatomist and keen socialist. With the help of his wife Cécile and a few other close collaborators, Vogt had been the first to map the cerebral cortex according to its cellular anatomy.

Faced with Lenin's rather shrunken brain, Vogt was unable to identify any particular region that suggested extraordinary mental prowess. But he nonetheless declared Lenin a cognitive "athlete", blessed with a heightened ability to make connections in the brain due to the unusual density of a certain type of cell in the upper layers of his cortex.

Although neuroscientists no longer believe that broad concepts such as genius can be pinned down to precise brain locations, they do want to know where more clearly defined cognitive activities take place. Today, they approach this task armed with sophisticated machines for picturing the structure and activity of living brains. Yet they still relate the resulting images to an anatomy mapped out in 1909 by Vogt's Berlin collaborator Korbinian Brodmann¹.

This is a serious limitation. Brodmann's map is at best imprecise and at worst downright wrong, says Karl Zilles of the C. & O. Vogt Brain Research Institute at the University of Düsseldorf in Germany, whose team is now trying to produce a brain atlas for the twenty-first century. "Neuroscientists obviously need a more usable map," he argues.

Brodmann examined a handful of human brains, scrutinizing them under a light microscope to divide the cerebral cortex into around 50 numbered areas that he judged to be anatomically distinct. That was a formidable achievement for the time, but today the map is of limited use in interpreting the detailed functional and structural images produced by magnetic resonance imaging (MRI) and positron-emission tomography. "The ultimate goal is to ascribe structure to function and vice versa," says Ray Dolan, who heads the Functional Imaging Laboratory at University College London. Without a new map, this task remains largely a matter of guesswork.



Spot the difference: Korbinian Brodmann's 1909 atlas (below left) does not account for variation between brains. In the composite image above, areas of high variation are shown in orange, yellow and green.

For one thing, Brodmann's two-dimensional map represents only the surface of the cortex, whereas structural or functional brain imaging gathers data in three dimensions. What's more, the exposed surface of the cortex represents only one-third of the total surface — Brodmann did not map the other two-thirds, which are hidden in deep folds called sulci.

In an attempt to address some of the problems with Brodmann's map, most brain imagers today refer to an atlas published in 1988 by Paris-based neuroscientists Jean Talairach, of René Descartes University, and Pierre Tournoux, of the Val de Grace Hospital². Talairach and Tournoux superimposed Brodmann's brain areas onto a three-dimensional description of one half, or hemisphere, of a single post-mortem brain. But this overlooks the fact that the brain is not symmetrical, and the map also overestimates the importance of sulci as landmarks of borders between distinct regions. "Worse still, it doesn't take into account the enormous variability in brain anatomy," says Zilles — neither

Brodmann nor Talairach and Tournoux realized that some brain areas vary in size by a factor of ten between different people.

Time to think

Over the past three decades, Zilles has developed a painstaking method for precision-mapping the cellular architecture of the human cortex. From the start, his goal has been to devise objective techniques that will not be biased by the neuroanatomist making the observations. "Brodmann's judgement about what constituted cortical areas was, by necessity, subjective — he could only look down his microscope and report," says Zilles, who is also head of the Institute of Medicine at the Jülich Research Centre, one of Germany's national laboratories.

Back in the 1970s, Zilles was viewed by many of his peers as an eccentric throwback to a bygone era. But today, the technology is in place to turn his vision into reality. The process begins at autopsy. For each brain being studied, Zilles's team makes an MRI scan before the organ is removed from the skull. They then embed the whole brain in a paraffin-wax block and cut it into slices just 20 micrometres thick — between 5,000 and 8,000 per brain.

Each slice is stored on a numbered glass slide, and the researchers stain every fifteenth one to display cell bodies. Every sixtieth slice is then taken for computer measurement of cell densities. This gives an interval of 1.2 millimetres between each analysed slide, which is better than the resolution of current functional MRI. And should brain-imaging technology improve, says Zilles, it will be possible to return to the unexamined slides and boost the resolution of the map. Thus, as the brain's functions are pinned down more precisely, knowledge of brain architecture can expand in tandem.

By entrusting the analysis to computers, Zilles has removed human bias from the business of defining borders between brain regions. First, images of each slide are compared with the corresponding section of the original MRI scan to show how they have been distorted by their preparation. After the tissue is 'morphed' back to the size and shape that it had *in situ*, the images are then subdivided into 20- μ m grids, and the computer determines the proportion of stained cell bodies to unstained tissue in each square to calculate its 'grey-level index'. Finally, the computer maps borders between distinct anatomical areas by looking for sudden and statistically significant changes in these indices. "You can't rely on visual inspection to see where the boundaries are between different cortical areas," says Zilles. "You have to let the computer define the edges."

Universal atlas

The final product will involve 15 brains to generate a 'probabilistic map', against which any point selected from, say, an MRI image of a living brain can be said to lie within a particular structure with a certain probability. This tackles the issue of variation in brain anatomy between individuals.

Even with a dozen or so staff on the job, the atlas will take years to complete. But Zilles has already collaborated with brain imagers to identify several new cortical areas, which the researchers have defined both functionally and anatomically. These include two areas known as the ventral and anterior intraparietal areas (VIP and AIP), which had been identified in monkeys but were thought to have no human counterpart.

In monkeys, the VIP is associated with perception of motion independent of the type of sensory input — whether the moving object is seen, heard or felt. The monkey AIP, meanwhile, recognizes three-dimensional objects, also independently of the type of sensory input. In MRI studies of humans, the brain regions that lit up in response to each of these functions corresponded to distinct anatomical areas in Zilles' evolving brain atlas^{3,4}.

Other neuroscientists are now collaborating with Zilles to try to pin down the anatomical locations of complex brain functions. For example, Per Roland of the Karolinska Institute in Stockholm, Sweden, has localized



Head on the chopping block: Karl Zilles (below) carves wax-encased brains into ultrathin slices before using a computer to map their structure.



different components of the processing that takes place when we explore an object by touch, such as the detection of changes in curvature, to precise areas of the cerebral cortex⁵.

More recently, Zilles has begun to examine the distribution of receptors for different neurotransmitters, the chemical signals by which brain cells communicate with one another. Using chemicals that bind to specific receptors, he has found that distinct anatomical brain areas also have characteristic 'fingerprints' of receptor density⁶. This is particularly exciting, says Dolan, as a combination of cellular architecture and signalling chemistry may provide important clues as to how brain functions are regulated. "The functional-imaging field is also moving in this direction," he says.

Zilles is also a partner in the International Consortium for Brain Mapping (ICBM), based at the University of California, Los Angeles (UCLA), which is producing a database of thousands of structural images of brains from healthy volunteers as a resource for neuroscientists. "Zilles' accurate, fine

anatomy will provide essential information to users," says UCLA neurologist John Mazziotta, one of the ICBM's principal investigators.

Another ICBM group based at UCLA, headed by Arthur Toga and Jacopo Annese, is working to derive a similar map. They are staining sections of ten brains for myelin, the fatty substance that sheaths nerve fibres, rather than the cell bodies themselves. "It will complement the Zilles map," says Annese. "It will show the connections between all of the neurons, which is also very important for understanding function."

When such atlases are finished, brain imagers should be in a better position to interpret their data. "MRI is a poor camera for what is going on in the brain, but it's the best we can do with living brains," comments Alan Evans of the Montreal Neurological Institute at McGill University in Canada, who was one of the founding members of the ICBM. Together with detailed anatomical atlases, it should become a more powerful tool.

Such a resource would undoubtedly have benefited Vogt when studying Lenin's brain. But it would have had its drawbacks too — Vogt might have found it rather more difficult to tell the Soviet authorities what they wanted to hear about their hero's cognitive abilities. ■

Alison Abbott is Nature's senior European correspondent

1. Brodmann, K. *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues* (J. A. Barth, Leipzig, 1909).
2. Talairach, J. & Tournoux, P. *Co-planar Stereotaxic Atlas of the Human Brain* (Thieme Medical, New York, 1988).
3. Bremner, F. et al. *Neuron* **29**, 287–296 (2001).
4. Grefkes, C., Weiss, P. H., Zilles, K. & Rink, G. R. *Neuron* **35**, 173–184 (2002).
5. Bodegård, A., Geyer, S., Grefkes, C., Zilles, K. & Roland, P. E. *Neuron* **31**, 317–328 (2001).
6. Zilles, K. et al. *Eur. Neuropsychopharmacol.* **12**, 587–599 (2002).

International Consortium for Brain Mapping

♦ www.loni.ucla.edu/ICBM

Virtual solution to carbon cost of conferences

Improved technology is making virtual meetings more like real ones, without the flights.

Sir — Every year, many thousands of scientists jet off to a host of destinations all around the world to attend conferences. Emissions of carbon dioxide, a greenhouse gas, from air travel are growing at an alarming rate. In 1992, at 500 million tonnes, they constituted about 13% of all CO₂ emissions from transportation sources. By 2050, the Intergovernmental Panel on Climate Change estimates that aircraft emissions will triple to 1.5 billion tonnes of CO₂ per year.

During the past few years, an increasing number of conference organizers have recognized this pollution problem and have attempted to mitigate their climate impact (see, for example, www.fhio.gc.ca/commuting/carbon_neutral.htm#Conf_guide). Some recent environmental conferences have aimed to offset the greenhouse-gas emissions caused by their participants' travel by funding tree-planting schemes. Others strive to be entirely 'carbon neutral', sourcing their electricity from renewable sources, and others even buy carbon credits to offset conference-related emissions. Now, a new technology, the Access Grid system (www.accessgrid.org), is promising to change the face of our conferencing habits.

Access Grid is similar to video

conferencing but lets groups from numerous different locations communicate among themselves at one time. Instead of only being able to see the speaker (a limitation of traditional video conferencing), delegates can also see and talk to other groups of delegates — in the next town, in another country or on a different continent. Speakers can manipulate their presentations to all viewers simultaneously so that, as each one moves through the slides on his or her own screen, all the viewers' screens are also updated.

A key aim of Access Grid is to solve the problems faced by many researchers in the developing world, who are prevented from attending international conferences by economic constraints. There is, however, a significant start-up cost of around US\$25,000 for the technology, and few locations in the developing world are currently able to meet the requirement of Access Grid conferences for broadband network access (greater than 1Mb per second minimum bandwidth).

Several international conferences have already successfully used the Access Grid, with many more such meetings planned and an ever-growing number of institutions able to co-host conferences. The huge environmental dividend of

virtual conferencing is demonstrated by the estimate for a recent genomics meeting, where travel-related CO₂ emissions of the order of 900 tonnes were avoided (www.ndsu.nodak.edu/virtual-genomics/Proc_VCGB2002.pdf, or see W. A. Valdivia-Granda, E. L. Deckard & W. Perrizo *Proc. Virt. Conf. Genom. Bioinf.* **1**, 1–3; 2002).

The combination of this new technology with ever-improving video and sound (for example, Hewlett-Packard's new Coliseum immersive teleconferencing system) means that the old objection — that virtual conferences are impractical and impersonal — is rapidly breaking down. Though many of us may feel it is unfair to deprive ourselves of all-expenses-paid international trips and the outside-meeting socializing common to most conferences, we should not ignore the environmental impact of these meetings. 'Real world' conferences will always have a place, but given the huge number of international conferences, even limited use of Access Grid virtual conferencing has the potential to reduce CO₂ emissions by many thousands of tonnes.

David S. Reay

*School of Geosciences, Ecology and Resource Management, University of Edinburgh, Darwin Building, Mayfield Road, Edinburgh EH9 3JU, UK
www.ghgonline.org*

Looking into the safety of AAV vectors

Sir — The News story "Harmful potential of viral vectors fuels doubts over gene therapy" (*Nature* **423**, 573–574; 2003) suggests that there is a reasonable probability that recombinant AAV vectors may cause or contribute to cancer in gene therapy subjects. As authors of the paper discussed in your story (H. Nakai *et al.* *Nature Genet.* **34**, 297–302; 2003), we would like to emphasize that there is no evidence that AAV causes cancer in animals or humans, and that your concern is unfounded.

Our studies demonstrate that AAV vector DNA will preferentially integrate into active genes when delivered into the livers of mice. This has raised concerns because of recently published reports of leukaemia in two of nine patients treated with a recombinant retroviral vector for a lethal genetic disease, X-linked severe combined immunodeficiency disorder (SCID). The leukaemia was caused at least in part by the retroviral insertion and activation of an oncogene (insertional mutagenesis) in

bone-marrow progenitor cells. Because retroviral vectors preferentially integrate into intragenic regions of the chromosome, your News story suggests that recombinant AAV vectors may pose similar risks in gene-therapy trials.

There are substantial differences between retroviral and AAV-mediated integration. First, unlike retroviral vectors, AAV-mediated vector integration is relatively uncommon. Second, retroviral vectors contain additional regulatory elements that are more likely than AAV vectors to activate a gene that they insert next to. Third, retroviral vectors contain the protein machinery needed to cause host chromosomal DNA breaks, whereas AAV does not. It is possible that AAV preferentially integrates into DNA regions that are already damaged within treated cells.

A symposium entitled "Safety considerations in the use of AAV vectors in gene transfer clinical trials", jointly sponsored by the NIH and the FDA, was held in March 2001 (see www4.od.nih.gov/oba/rac/Transcript3-7-011.pdf). On the basis of data from hundreds of normal mice treated with the vector, the conclusion was

reached that there was no evidence to suggest that the vector caused cancer.

In addition, the leukaemia found in patients treated with X-linked SCID gene therapy may be unique to this particular disease because of the unusual physiological events that occur after treatment. In X-linked SCID, the genetic reconstitution of a very few precursor cells results in the selective proliferation of immune cells genetically corrected with the vector. Any additional proliferation stimulus, such as the activation of an oncogene, may result in the further growth and expansion of these cells. This type of growth advantage is not a factor in most gene-therapy trials, and AAV has not been used in clinical trials to treat such disorders.

Although we support additional long-term safety studies, we believe the risk of cancer in current AAV trials is negligible, on the basis of infrequent integration efficiency and the quiescent nature of the target tissues.

Mark A. Kay, Hiroyuki Nakai

Departments of Pediatrics and Genetics, Program in Human Gene Therapy, Stanford University, 300 Pasteur Drive, Stanford, California 94305-5208, USA

A troubled pilgrim's progress

The compelling personal journey of a founding father of cellular immunology.

Science as Autobiography: The Troubled Life of Niels Jerne

by Thomas Söderqvist

Yale University Press: 2003. 400 pp.

\$40, £27.50

Gustav J. V. Nossal

The central puzzle of the immune system is how it can react specifically to virtually any microbe or foreign molecule. How is it possible, with a large but limited number of genes, to produce billions of different sorts of antibody? Before revealing the answer, we must pay tribute to Niels Kaj Jerne (1911–94), the Danish-born immunologist who unquestionably deserves the credit for creating a theoretical framework within which the matter could be pursued.

The steps along the path are elegantly described in this biography by Thomas Söderqvist, first published in Danish in 1998 and recently revised and translated into English. But if the reader is expecting a typical scientific biography, in which the research of a famous person is catalogued and explained, a rude shock awaits. After more than ten years of meticulous examination of Jerne's extensive papers, and 160 hours of interviews with him, Söderqvist presents a thought-provoking, surprisingly frank account of Jerne's personal life and philosophy.

The controversial justification for this approach is Söderqvist's belief that the source of Jerne's creativity lay in his personal experiences; that his scientific work was an inseparable part of his life as a whole. Thus, Söderqvist embarks on what he terms new biographical territory, namely existential biography. As a result, we are given both an authoritative presentation of Jerne's scientific achievements and a penetrating survey of this troubled pilgrim's progress.

Jerne's parents were Danish, but lived successively in England, Germany and the Netherlands. The young Niels became fluent in Danish, English, German and Dutch, never feeling rooted anywhere. He gradually assumed the mantle of an élitist, epicurean, deeply cultured European intellectual. A 'bookworm' from a very young age, he steeped himself in Kierkegaard, Bergson, Nietzsche, Proust and Gide. Paradoxically, he felt himself to be insecure, alienated and a misfit while at the same time enjoying conversation, longing for the sublime and, convinced of his superior gifts, eager to make an impression on the world.

He took a long time to find his *métier*. After school, he worked in business in Rotterdam, then moved to Leiden to study mathematics and physics, dropping out after



Source of inspiration? Niels Jerne's personal life may have contributed to his scientific work.

a couple of years. At 22, he decided to study medicine in Copenhagen, but again could not commit himself to it. He failed in a job with a publishing house, then worked for his father for three years — helping him to find better ways of curing bacon. Only at the age of 28 did he begin to study medicine seriously, graduating eight years later, and starting research at the State Serum Institute in Copenhagen.

At 23, Jerne had married his first wife, Tjek Wahl, a German art student who established a successful painting career and bore him two sons. Marriage problems began seven years later, when Jerne had a two-year affair with a physician whose erotic arts included sado-masochism. Although Tjek knew of this and of two other brief liaisons, she was unable to accept it when Niels started an affair with her close friend Adda.

As Tjek's own infidelities became clear, Niels believed divorce to be the only option. But after ten years of marriage, Tjek could not bear this thought and committed suicide. Jerne never forgave himself. In time he married Adda, who was soon reduced to the role of stepmother and housekeeper. In his mid-forties, Jerne began an affair with a 19-year-old girl who (after divorce from Adda) became his third wife. Adda in a relationship with an American woman who bore his third child, and one sees the complexity of this man's private life.

Jerne's first major scientific achievement was in formulating the natural-selection theory of antibody formation, published in 1955. He postulated that millions of different

kinds of antibody are synthesized naturally, without previous exposure to foreign antigens, and that their diversity is due to a random process. The antigen, rather than being involved in moulding the antibody, is only a catalyst. David Talmage and Frank Macfarlane Burnet refined the theory, proposing that the natural antibody is a surface receptor on white blood cells known as lymphocytes, with the antigen needing only to encounter the 'right' cell to stimulate the immune response and increase the number of appropriate cells. This clonal-selection theory turned out to be correct. Regarding the antigen as playing a selective rather than an instructive role was a true paradigm shift.

Jerne made two other profound contributions. He devised a simple and elegant technique (known as the haemolytic plaque technique) for determining the number of antibody-forming cells. This added a vital quantitative element to cellular immunology, and facilitated the development of hybridoma technology for producing monoclonal antibodies — antibodies with predetermined exquisite specificity. Jerne shared the 1984 Nobel prize with its inventors, Georges Köhler and César Milstein. He was also the founding director of the Basel Institute for Immunology which, within a decade, he made into a Mecca of European immunology and a cauldron of intellectual ferment. It must have been a source of great satisfaction to Jerne that it was at this institute that the nature of the generator of random antibody diversity was discovered by Susumo Tonegawa, leading to another Nobel prize. Each antibody-forming cell develops from a unique recombination event that assembles functional antibody genes from a pre-existing array of minigenes. The combinatorial element of antibody-gene production thereby creates the huge number of different end products.

Jerne also deserves credit for setting up an immunology unit within the World Health Organization to promote research and teaching in developing countries. But his network theory of immune responses proved less successful, and his impatience with experimental details lessened the impact that he might otherwise have had on the development of immunology as the science reached maturity. As Söderqvist remarks: "Jerne never took satisfaction from small discoveries — in this respect he resembled an artist more than a biomedical researcher — and he struggled to formulate theories of great breadth, wishing to impose his own worldview and his personality on nature." Later in life, Jerne complained that

immunology was becoming too complicated and technical. His dismissal of the reductionist approach to the discipline prevented him from appreciating the amazing discoveries made in immunology during the last 15 years of his life.

Is Söderqvist successful in his attempt at existential biography? The answer must be hedged in uncertainty. Did Jerne use the experiences of his inner life, his understanding of himself, to construct the natural-selection theory? It seems far-fetched, but who has the right to be dogmatic about the wellsprings of human creativity? Where Söderqvist has undoubtedly succeeded is in revealing the startling personal journey of one of the twentieth century's great biological thinkers, and in sketching the milieu within which cellular immunology came of age. This serious work of scholarship will be devoured both by immunologists and by a wide general readership. ■

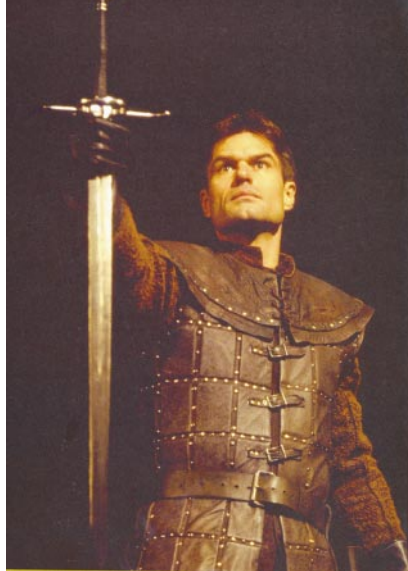
Gustav J. V. Nossal is in the Department of Pathology, University of Melbourne, Victoria 3010, Australia.

Scanning Shakespeare

The Bard on the Brain: Understanding the Mind Through the Art of Shakespeare and the Science of Brain Imaging
by Paul M. Matthews & Jeffery McQuain
Dana Press: 2003. 192 pp. £24.50
University of Chicago Press: 2003. 192 pp. \$35
Kevan A. C. Martin

Every now and then, an editor will come up with what seems to them to be a brilliantly original wheeze. Here, Jane Nevins of the Dana Press persuaded Paul Matthews, a neurologist, and Jeffery McQuain, an English scholar, to use Shakespeare's plays as a vehicle for describing the brain correlates of cognitive function. Perhaps she was prompted by the popular success of the 1998 film *Shakespeare in Love*, in which the audience meets the young Shakespeare, then struggling with writer's block and trying to meet a deadline for his new play, *Romeo and Ethel the Pirate's Daughter*, whose plot is suggested to him by his rival, Christopher Marlowe. More significantly, the film conveys the powerful effect of a live theatre performance on an audience. How does theatre weave its magic spell? In the film, Phillip Henslowe, director of The Rose Theatre, gives his answer: "I don't know. It's a mystery."

The connection between theatre and the brain explored in this book is less mysterious when we learn that the remit of the Dana Alliance for Brain Initiatives is to publicize



"Once more unto the breach": Shakespeare's Henry V used mental imagery to rally his troops.

information about the benefits of brain research. The most visible and successful of the initiatives is Brain Awareness Week, a popular event held annually in many centres worldwide. It provides an invaluable forum for members of the public to meet patient-support organizations and local brain scientists, who usually cover a wide breadth of basic and clinical research.

Square and 'coffee table' in its aspect, there is much that is artful in this book. It is itself a play in seven 'acts'. Each act covers different topics, such as "Minds and Brains", "Our Inner World" and "The Seventh Age of Man", and each contains a number of 'scenes', whose titles range from the straightforwardly banal, "The Wonder of the Human Brain", to the dramatic, "Let Me Clutch Thee", to the boggling, "Putting an English Tongue in a French Brain". Each scene begins with a précis of a plot or subplot from one of the plays, which is followed by an excerpt from a scene and finally a commentary that interweaves topics evoked by the excerpt and its speculative neurological correlate. With 34 such scenes, this unrelenting format does wear a little thin. Writing anything worth reading alongside an excerpt from Shakespeare is a challenge, but here even the shortest sentences can somehow contrive to emphasize the difference, as in "Shakespeare has Juliet ponder what defines a 'Montague'".

The subsequent analyses are largely un-revealing: "Shakespeare was a keen observer of human nature", or "some of his characters talk to themselves in sonnets". Unfortunately, much of the commentary on Shakespeare's text is reminiscent of the cribs one crammed the night before the English literature examination in the hope of impressing the teacher with such pearls as "Shakespeare lived intimately with a rich world of imagination, which he communicated to others through words and stage action".

The book's illustrations are prodigal. The full-page photographs of performances of Shakespeare's plays appear without context

or connection to the text. The neurological illustrations are small, 'arty' graphics of the computer-generated variety, a technique that has too easily enabled images of the brain, each postage-stamp-sized, to multiply on obscure and complicated backgrounds. Again, the repetition of style makes these brain images interchangeable. What's actually being illustrated is not only hard to see, but is poorly explained, despite lengthy captions, albeit of variable fidelity. The caption of figure 9, for example, implying that the right and left visual fields map onto the primary visual cortex of the same hemisphere, could be written off as a schoolboy howler, were it not that the accompanying text confirms that the authors really have misinterpreted the experiment illustrated, which shows the distribution of cortical visual areas and not, as they state, the well-known 'ocular dominance columns' of the primary visual cortex, which are beyond the spatial resolution of the scanners used.

The neurological material is restricted mainly to functional imaging of the human brain, principally the "folded surface of the brain (called the cortex) ... where neurons are found". Unfortunately, the different imaging methods — for example, positron-emission tomography (PET) and functional magnetic resonance imaging (fMRI) — used in the various studies referred to in the book are not properly explained. There is no doubt that fMRI has revolutionized cognitive neuroscience. The method, however, does not actually detect "the amount of fresh blood flow", but records changes in the levels of the paramagnetic molecule deoxyhaemoglobin. How this blood-oxygen-level-dependent signal (the so-called BOLD signal) correlates with neural activity has been the subject of intense and technically difficult invasive experiments with animals, but such experiments are vital steps in developing a deeper understanding of the basic physiology and anatomy underlying human brain function.

This information would help lay readers to understand why the non-invasive techniques now used in human studies cannot, on their own, "define the brain mechanisms responsible for thoughts, emotions, and disorders that Shakespeare wonderfully described". The reader is left with the impression that imaging techniques all produce multiple images of tiny brains decorated with rainbow-coloured blobs of unknown significance. This is a missed opportunity to explain why different techniques of brain imaging are used in the clinical setting, and how the images are interpreted.

A danger inherent in popular science books is that they try to be popular and so simplify too much. The lesson we learn from the theatre, and from Shakespeare's plays in particular, is that complex ideas and emotions can be effectively communicated to a receptive audience, even if the language is

unfamiliar. The large audiences at Brain Awareness Week indicate overwhelmingly their eagerness to learn about the brain, but if these audiences leave with the impression that scientists have discovered that communication between humans arises from coloured blobs in our left hemispheres, then we have missed the essential humanity of brain sciences. Encouragingly, the dialogue between brain researchers and their lay audiences is growing, thanks in good measure to the Dana Alliance, so, strangely enough, things will probably turn out well. How? I don't know. It's a mystery. ■

Kevan A. Martin is at the Institute of Neuroinformatics, University of Zurich/ETH, Winterthurerstrasse 190, 8057 Zurich, Switzerland.

There and back again

The Development of Animal Form: Ontogeny, Morphology, and Evolution

by Alessandro Minelli
Cambridge University Press: 2003. 342 pp.
£55, \$75

Axel Meyer

"There and back again", the alternative title of J. R. R. Tolkien's *The Hobbit*, is a fitting way to describe the intellectual journey that every generation of self-respecting biologists has travelled since Ernst Haeckel, Karl Ernst, Ritter von Baer, Georges Cuvier and Johann Wolfgang von Goethe before them. What they and, in cycles of 20 years or so, intellectual giants such as John Tyler Bonner and Stephen Jay Gould have tried to understand is the interwoven relationship between development and evolution. Every new generation of comparative biologists goes 'there' and advances the field further, building on the insights of the previous generation's work by applying new methods and techniques. But much of what had been thought about a generation before is often forgotten and, importantly, new questions are raised as well. So, every generation of biologist goes 'back again' to this big issue.

Selection can only act on things that are developmentally possible. In other words, developmental mechanisms constrain evolutionary possibilities, and they are often very conservative, carrying the "load" of previous evolutionary lineages, as Alessandro Minelli puts it. Sometimes, however, as in the case of the direct and indirect development (without and with free-swimming larvae, respectively) of closely related species of sea urchin, development can also be surprisingly variable. But how do developmental mechanisms themselves change during evolution, and how does evolution in turn affect

Portrait Updating Hooke

In an immaculate seventeenth-century interior reminiscent of Jan Vermeer's painted rooms stands Robert Hooke, who died 300 years ago. He is equipped with a set of instruments, including the microscope needed for his *Micrographia*. Isaac Newton, whom some say is responsible for the loss of the only authentic painting of Hooke, lurks outside the window. Glancing out uneasily, Hooke teasingly spins his globe. This ingeniously contrived 'photographic' image (right) is Guy Heyden's winning entry in the competition "Portraying Robert Hooke – Recreating the Hidden Genius". Heyden carries off the £500 prize, awarded by the Royal Institution of Chartered Surveyors and the Royal Society. The brief was to create a "replacement" portrait, not as a recreation of the lost picture but in a twenty-first-century style. Just as Vermeer used the high-tech of his day, a camera obscura, so Heyden has used a computer to create a 'reality' analogous to the optical realism of the paintings known to Hooke. **Martin Kemp**



development? Every generation of biologists during the past 180 years or so has made progress on the quest for a deeper understanding of these questions, through both the conceptualization and the application of the new techniques of their time.

Gould's 1977 historical treatise *Ontogeny and Phylogeny* (Harvard University Press) was probably the most influential contribution to this field in the previous generation. In its current reincarnation the field now has a name, evolutionary developmental biology, or evo-devo, and is influenced by the thinking and research of a new range of scientists, including Minelli.

What new ideas have emerged in the past 10–15 years? In a nutshell, the comparative application of molecular developmental methods that are interpreted in a rigorous (often molecular) phylogenetic framework. Recent comparative developmental and genomic studies have yielded the apparently paradoxical insight that many genes (particularly Hox genes) and their interactions in genetic networks are astonishingly conserved in evolution. These results were unexpected and raised the question of how the diversity of body architecture in different phyla has arisen, given that genetically so much has remained the same during the past several hundred million years. The historically static view of homology proposed in the nineteenth century by Richard Owen, and still widely taught today, has also been revolutionized by these comparative developmental studies and, in my opinion, has been largely abolished. Subsequent phylogeny-based theories of homology are increasingly being questioned because it is unclear whether developmental processes and

mechanisms should be part of the definition of homology. Minelli carefully dissects the concepts of absolute and relative or partial homology into its components and discusses them from an evo-devo perspective.

'Renaissance man' is a term often used for a member of the (now all too rare) breed of scientists who have a wide range of intellectual interests — those who are still real scholars. This seems a very appropriate descriptor for Minelli. If you want to make significant advances in the ancient evo-devo field you need to be able to see the big picture and — listen up students — know the old literature. Minelli's daily bread-and-butter research deals with the taxonomy, systematics and comparative morphology of arthropods, particularly the myriapods, which have many segments. But he has long been an important contributor to evo-devo with his conceptual work on segmentation, modules, homology, appendages and body axes, all of which is founded on an impressive knowledge of biodiversity and comparative morphology. This impressively scholarly book summarizes and further develops his work of the past 30 years on basic features of the organization of animals, such as body axes, symmetry, segments, appendages and homology.

Rather than asking small questions, Minelli presents insightful hypotheses and concepts. If you want to look up from your myopic concentration on your single model system and broaden your horizons, you should read this book. It is a 'must read' for any practitioner in the fields of developmental and evolutionary biology — fields that, at long last, are beginning to be unified. ■

Axel Meyer is in the Department of Biology, University of Konstanz, 78457 Konstanz, Germany.

In the spotlight

Summarize yourself in the form of a title of a paper in Nature.

Geneticist eavesdrops on bacterial conversations and learns to translate multiple chemical languages.

What was your first experiment as a child?

I tried to make a candle-propelled steamboat based on a magazine article for kids. It was a total flop — sank like a rock, which partly explains why I did not end up in physics or engineering.

What makes a good scientific mentor?

Patience, a good sense of humour, and always encouraging the student to believe that he or she will figure it out.

What single scientific paper or talk changed your career path?

A talk by Michael Silverman about how to use luminescent bacteria as reporters for cell–cell communication. I heard him talk and asked him that day if I could be his postdoc.

What book has been most influential in your scientific career?

Lehninger's *Principles of Biochemistry*. There was a time when I virtually had it memorized.

What gives you the most job satisfaction now?

What are your major frustrations?

Satisfaction: when students or postdocs tell me about some wonderful, creative experiment or result that they came up with on their own. Frustration: trying to get grants to fund the work.

What's your favourite conference destination, and why?

Rome. Great history, sites, museums, food and climate.

What was the worst/most memorable comment you ever received from a referee?

"Genetic experiments mean nothing."

What book is currently on your bedside table?

Alessandro Baricco's *Ocean Sea*.

What music heads the playlist in your car or lab?

Kate Rusby's *Hourglass*.

Assuming the dead can be raised and/or time travel exists, who from the world outside science would you most like to have dinner with?

Virginia Woolf. Her book *A Room of One's Own* had a tremendous impact on me when I was younger. I think she would have a lot to say, especially about being true to one's self, finding one's passion, and carving out a space and place for oneself.

Under what conditions do you have your greatest and most inspired ideas?

Standing in a hot shower with the water pouring over me.

Where and when would you most like to have lived or worked?

I'm content as is.

You are on a plane behind two students obviously going to the same conference, who start to talk about your work. What do you do?

Listen in of course. That's what I do for my work anyway. Bacteria ... people ... it's all the same to me.

What one thing would you rescue from your burning laboratory?

After all the people, our bacterial strain list.

What's the best piece of advice you've ever received?

"Bonnie, the people who need to know, know."

What do you most dislike about having research published?

It takes us years to think up screens, generate mutants, and identify what the genes do and how the components interact. Once we publish, we have to give the strains out to everyone who asks.

What's the most interesting thing in your fridge?

Eleven different kinds of mustard.

Why is physics so hard?

Because physicists have a pact with each other to speak in gibberish. The sole point is to keep biologists from learning how easy physics and maths really are. That's my theory anyway, or it could be that biologists are bad at disciplines that involve adhering to rules and expressing oneself in mathematical terms.

The Internet is the bane of scientists' lives because...

...e-mail is so distracting.

What do you do to relax?

Singing and acting lessons, aerobics class, hiking, bike riding, reading and cooking.

What would you have become, if not a scientist?

I would love to be an actress in the theatre.

What previously under-recognized sport or pastime should be included in the Olympic Games?

Ballroom dancing.

What single discovery, invention or innovation would most improve your life?

A self-cleaning house and self-cooking meals.



Bonnie Bassler

Bonnie Bassler is a professor in the molecular-biology department at Princeton University. She teaches aerobics, and takes singing, dancing and acting lessons. Her hobbies include reading, cooking and seeing live theatre.

What's the one thing about science that you wish the public understood better?

That we are not nerds, but artists. That what we do and what we are is exciting, creative and fun.

Do you have a burning ambition to do or learn something of no practical or immediate value?

I would be a contortionist in the Cirque du Soleil.

Which field in science (apart from your own) deserves more funding, and why?

The interface of theoretical physics/mathematics and biology.

What music would you have played at your funeral?

Bach's unaccompanied cello suites (Prelude from Suite Number 1 in G Major, BWV 1007).

How would you like to be remembered?

That I was creative, passionate and authentic; that I always loved science and was always willing to be surprised; that I was a clear communicator; that my group always did first-rate science, and that I always gave the credit for the work to the students and postdocs who did it.

What's your motto?

You make your luck.

You are fond of theatre: which actor would best portray you in a film?

Lily Tomlin.

What's just around the corner?

Even more surprises. Proudly waiting, with bittersweet anticipation, to be put out of business by my students. ■

Evolution in population dynamics

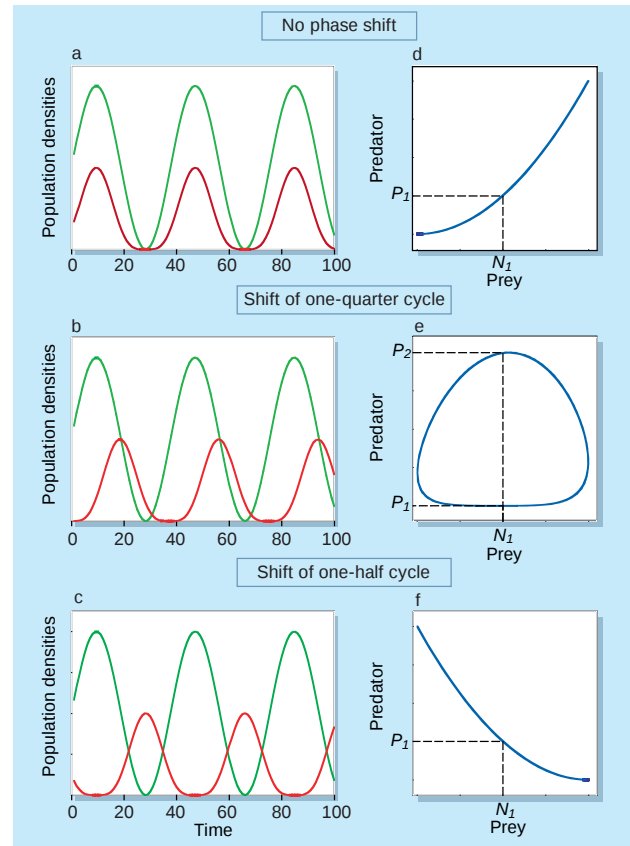
Peter Turchin

In their study of predator–prey cycles, investigators have assumed that they do not need to worry about evolution. The discovery of population cycles driven by evolutionary factors will change that view.

Ecologists studying population dynamics prefer not to bother with the possibility of evolutionary change affecting their study organisms. This is sensible, because understanding the results of interactions between, for example, populations of predators and prey is already a complicated task. Making the assumption that evolutionary processes are too slow on ecological scales greatly eases the task of modelling the commonly observed population oscillations. But an elegant study by Yoshida *et al.*¹ (page 303 of this issue) decisively demonstrates that this simplification might no longer be tenable.

The story begins at Cornell University when two members of the group — ecologist Nelson Hairston Jr and theoretician Stephen Ellner — teamed up to study the population dynamics of rotifers (Fig. 1), microscopic aquatic animals that feed on unicellular green algae. According to ecological theory, the interaction between predators (such as rotifers) and prey (algae) has an inherent propensity to oscillate². Predators eat prey and multiply, causing prey numbers to crash, which in turn leads to a decline in the starving predator population, allowing prey to increase, and so on. Indeed, when the investigators placed rotifers and algae in a ‘chemostat’ (a laboratory set-up with continuous inflow of nutrients and outflow of waste) they observed population cycles³. But the phase shift between predator and prey cycles was completely ‘wrong’ — predators peaked when prey were at the minimum and vice

versa, resulting in almost perfectly out-of-phase oscillations. This is a subtle but important point, which requires an explanation. Suppose we observe three ecosystems containing predators and their prey. These three systems are in all ways identical, except in the phase shift between predators and prey: no shift (Fig. 2a), a shift of one-quarter of a cycle (Fig. 2b), and a shift of half a cycle (Fig. 2c). Clearly, there is some sort of dynamical connection between the two populations in all cases, but in which case are cycles driven by the predator–prey interaction? To answer this question we replot each trajectory in the ‘phase space’ — two-dimensional euclidean space in which each variable (prey and predator density) is represented with its own axis (Fig. 2d–f). When oscillations are synchronous, the trajectory goes back and forth along the same path, so that for each value of prey density (say, N_1) there is just one corresponding value of predator density (P_1). This means that if we already know the level at which



versa, resulting in almost perfectly out-of-phase oscillations. This is a subtle but important point, which requires an explanation.

Suppose we observe three ecosystems containing predators and their prey. These three systems are in all ways identical, except in the phase shift between predators and prey: no shift (Fig. 2a), a shift of one-quarter of a cycle (Fig. 2b), and a shift of half a cycle (Fig. 2c). Clearly, there is some sort of dynamical connection between the two populations in all cases, but in which case are cycles driven by the predator–prey interaction? To answer this question we replot each trajectory in the ‘phase space’ — two-dimensional euclidean space in which each variable (prey and predator density) is represented with its own axis (Fig. 2d–f). When oscillations are synchronous, the trajectory goes back and forth along the same path, so that for each value of prey density (say, N_1) there is just one corresponding value of predator density (P_1). This means that if we already know the level at which

prey is, knowledge of predator numbers gives us no additional information. In a differential equation describing prey dynamics we can replace all terms containing P with N by using the relationship depicted in Fig. 2d, leaving us with a single differential equation for N . But mathematical theory tells us that such single-equation models cannot generate cycles⁴. In other words, simply by noting that prey and predators oscillate in synchrony we have disproved the hypothesis that cycles are driven by the classical predator–prey mechanism described in the previous paragraph. Some other factor must be involved in producing the oscillations.

The same logic applies to the case of perfectly out-of-phase oscillations (Fig. 2c, f). However, in the case in which predators trail prey by a quarter of a cycle there are two values of P for each N (Fig. 2e). So a single equation for prey does not suffice; we must know what predators are doing. If predators are at the low point (P_1), prey will increase, but if



Figure 1 Predator: the rotifer *Brachionus calyciflorus*¹.

predator numbers are high (P_2), prey numbers will crash. The full model for the system will have two equations, one for prey and one for predators, and we know that such two-dimensional models are perfectly capable of displaying cyclic behaviour.

We now see why the observation of the perfectly out-of-phase dynamics demonstrates that the rotifer–algal cycles could not be driven by the classical predator–prey mechanism. So what is the actual explanation? The path taken by the Cornell group to answer this question is an almost textbook example of how science is supposed to be done. First they advanced four competing hypotheses, suggested by various known features of algal and rotifer biology. Next they translated the hypotheses into mathematical models and contrasted model predictions with data. Only one model, based on the ability of algae to evolve rapidly in response to predation, successfully matched such features of data as period lengths and phase relationships⁵. This is a convincing result, and if we dealt with a natural system we would have to stop there because we cannot usually manipulate the genetic structure of field populations.

In the laboratory, however, such an experiment is possible, and the successful test reported by Yoshida *et al.*¹ provides the final and most decisive evidence of the rapid-evolution hypothesis. Thus, the out-of-phase cycles result from the following sequence of observed events: under intense predation, the prey population becomes dominated by clones that are resistant to predators; when most prey are resistant, the predators remain at low numbers even though prey abundance recovers; low predation pressure allows non-resistant clones to outcompete resistant ones; so predators can increase again, leading to another cycle.

The experimental demonstration that rapid evolution can drive population cycles means that ecologists will have to rethink several assumptions. To give just one example, there is a long-standing debate in population ecology on whether natural populations can exhibit chaotic dynamics. Chaos (in the mathematical sense) is irregular dynamical behaviour that looks as though it is driven by external random factors, but in fact is a result of internal workings of the system. Before the discovery of chaos, ecologists thought that all irregularities in the observed population dynamics were due to external factors such as fluctuations of climate. Now we realize that population interactions (including those between predators and prey) can also result in erratic-looking — chaotic — dynamics. Incidentally, the chaos controversy was the main reason why the Cornell group decided to study rotifer population cycles.

Some ecologists have argued that chaotic dynamics cause populations to crash to very

low densities at which the probability of extinction is high, and that natural selection should therefore cause evolution away from chaos⁶. Since this argument was advanced, at least two examples of chaotic behaviour have been discovered: in the dynamics of the incidence of childhood diseases such as measles⁷, and of population numbers of rodents such as voles and lemmings⁸. What is more important, however, is that the argument assumes that evolution occurs on much longer timescales than oscillations. But the results of Yoshida *et al.*¹ show that evolution can be an intrinsic part of oscillations, raising an exciting possibility that some populations might rapidly evolve both towards and away from chaos. Perhaps this is the explanation of the puzzling observation that some Finnish vole populations shift from a stable regime to oscillations, whereas others do precisely the reverse⁹.

This is rank speculation, however, and will have to remain so because we cannot test it experimentally in natural systems. But in

the laboratory much more is possible, as the study by Yoshida *et al.* shows. We can hope that in the near future we will see an experimental investigation of the possibility of rapid evolution to and away from chaos. ■

Peter Turchin is in the Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, Connecticut 06269, USA.

e-mail: peter.turchin@uconn.edu

1. Yoshida, T., Jones, L. E., Ellner, S. P., Fussman, G. F. & Hairston, N. G. Jr *Nature* **424**, 303–306 (2003).
2. May, R. M. in *Theoretical Ecology: Principles and Applications* 2nd edn (ed. May, R. M.) 5–29 (Sinauer, Sunderland, Massachusetts, 1981).
3. Fussmann, G. F., Ellner, S. P., Shertzer, K. W. & Hairston, N. G. Jr *Science* **290**, 1358–1360 (2000).
4. Edelstein-Keshet, L. *Mathematical Models in Biology* (Random House, New York, 1988).
5. Shertzer, K. W., Ellner, S. P., Fussman, G. F. & Hairston, N. G. Jr *Anim. Ecol.* **71**, 802–815 (2002).
6. Berryman, A. A. & Millstein, J. A. *Trends Ecol. Evol.* **4**, 26–28 (1989).
7. Tidd, C. W., Olsen, L. F. & Schaffer, W. M. *Proc. R. Soc. Lond. B* **254**, 257–273 (1993).
8. Turchin, P. *Complex Population Dynamics: A Theoretical/Empirical Synthesis* (Princeton Univ. Press, 2003).
9. Hanski, I. *et al. Ecology* **82**, 1505–1520 (2001).

Accelerator physics

In the wake of success

Robert Bingham

Particle accelerators tend to be large and expensive. But an alternative technology, which could result in more compact, cheaper machines, is proving its viability for the acceleration of subatomic particles.

Since the construction of the first particle accelerator in 1932, high-energy collisions of accelerated ions or subatomic particles (such as electrons and their antimatter counterpart, positrons) have proved a useful tool in physics research. But the escalating size and cost of future machines means that new, more compact acceleration techniques are being sought. In *Physical Review Letters*, Blue *et al.*¹ report results from a test facility at the Stanford Linear Accelerator Center (SLAC), Califor-

nia, that have great significance for the future of particle accelerators. Their success heralds an entirely new type of technology, the plasma wake-field accelerator.

When charged particles such as electrons or positrons pass across a gradient of electric field, they are accelerated — how much depends on the steepness of the gradient. In conventional accelerators, a radiofrequency electric field is generated inside metal (often superconductor) accelerator cavities. But the gradient can be turned up only so far before



Figure 1 The wake created by a boat is a familiar image, but it is also the inspiration for a new type of particle accelerator. Blue *et al.*¹ have demonstrated that waves in a hot, ionized plasma of gas can create a rippling electric field in their wake, and that this 'wake field' can accelerate subatomic particles.

ALAN SCHEIN/CORBIS

the field breaks down. A more attractive medium for the next generation of particle accelerators is a plasma—a gas of atoms that has been ionized into a soup of electrons and nuclei. Plasmas can support electric fields with strengths of around a hundred gigavolts per metre, which is orders of magnitude higher than conventional accelerators.

At the centre of Blue and colleagues' apparatus¹ is a chamber 1.4 metres long containing a hot plasma created from lithium atoms. The electron density in the plasma is about 1.8×10^{14} per cubic centimetre. When a tightly packed bunch, or pulse, of positrons is fired into the plasma, it creates a rippling electric field behind it, rather like the wake of a boat (Fig. 1), and hence the name 'wake field'. Inside the chamber, the positively charged particles in the pulse 'suck in' negatively charged electrons from the plasma. To restore charge neutrality, the displaced electrons then snap back, away from the pulse, but tend to overshoot their original positions. This oscillation in plasma electron density creates an oscillating electric wake field, moving behind the pulse (Fig. 2).

Roughly the first two-thirds of the positron pulse (which overall comprises about 10^{10} particles) lose energy in setting up the intense plasma wake field. From their initial energy of 28.5 giga-electronvolts (GeV), these positrons decelerate at a rate of 49 mega-electronvolts (MeV) per metre. But the wake field they create accelerates the trailing third of the beam, increasing these particles' energies by almost 80 MeV over the length of the chamber, which corresponds to an accelerating gradient of 56 MeV per metre. The results are in excellent agreement with a three-dimensional simulation of the process, which predicted a peak energy gain of 78 MeV.

This is the first experiment to demonstrate energy gain by positrons in a plasma, and to use and accelerate particles that are at energies of interest for high-energy physics. (An existing electron–positron collider at SLAC used beams of particles with energies up to 50 GeV; the 'next linear collider', still at the planning stage, will have beam energies ultimately of 500 GeV.) To achieve this, Blue *et al.*¹ have had to overcome some tough practical problems. For example, the transportation and focusing of the positron pulse in the long plasma chamber is a remarkable achievement. The pulse shows no evidence of instabilities, particularly of the type known as 'hosing' instabilities², which induce transverse oscillations in the tail of the pulse that eventually destroy it. It is also significant that the positrons can propagate in the plasma without being annihilated by their electron antiparticles.

High-gradient acceleration of electrons as well as positrons is necessary for the development of a compact electron–positron plasma collider. In fact, the positron pulse

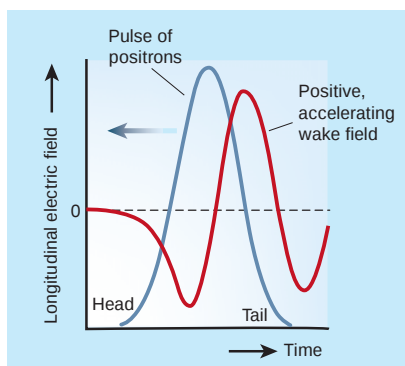


Figure 2 Beam-driven plasma acceleration. A pulse of positrons (blue line) is fired into a plasma of lithium gas. The positively charged particle pulse attracts the negatively charged electrons in the plasma, drawing them in and creating an electric field in the wake of the pulse (red curve). Initially the longitudinal field dips negative, but then the displaced electrons snap back, overshooting their original positions and driving the longitudinal field to positive values. This strong, oscillating field is termed a 'wake field'. Although particles at the head of the pulse lose energy in setting up the wake field, particles in the tail are accelerated by it.

used by Blue *et al.*¹ could simply be replaced by an electron pulse. The wake field would then arise through a slightly different mechanism: the electron bunch expels all the plasma electrons within its radius (known as 'blow-out'), leaving a channel of positively charged ions or nuclei behind it; again, the oscillating charge density generates a wake field that will accelerate the tail-end of the bunch. Calculations show that the wake fields generated by positrons are weaker than for electrons, but the two would be comparable if, rather than being uniform, the plasma were shaped to form a hollow cylinder³.

Instead of particles, a laser pulse can be used to set up an accelerating wake field in a plasma. In this case, the radiation pressure of the laser pushes the plasma electrons out of its path, creating oscillations in its wake.

Ageing

Microarraying mortality

David Gems and Joshua J. McElwee

Understanding how we grow old is a long-sought goal. A new large-scale study of gene expression in worms allows us to glimpse the complex biochemistry of lifespan.

The past decade has seen dramatic developments in studies of the genetics of ageing and longevity, mostly involving model organisms such as the nematode worm *Caenorhabditis elegans*, baker's yeast, fruitflies and mice¹. This has created considerable optimism that an

Previous experiments⁴ on plasma wake fields driven by terawatt lasers have shown that particles injected into the wake field gain about 200 MeV of energy, but the acceleration has only been sustained over millimetre distances; Blue and colleagues' beam-driven acceleration is sustained over more than a metre.

Perhaps the first incarnation of the plasma wake-field accelerator will be within a conventional accelerator as a 'plasma afterburner'⁵. By introducing 10-m plasma sections into a linear collider that is many kilometres long, wake-field acceleration could be used to double the energy of conventionally accelerated particles. But wake-field acceleration comes at the price of decelerating the head of the particle bunch. To counteract this, plasma lenses are included in the design, between the afterburner and the collision point. Such lenses focus the particles into a tighter bunch, far more strongly than do the conventional magnetic variety⁶. The tighter the bunches are focused, the more likely are particles in opposing bunches to collide when they meet, so the rate of collisions is not degraded by inclusion of the afterburner.

The work presented by Blue *et al.*¹ marks the most significant progress so far in the field of advanced accelerator research. It is likely to have an impact on many fields, as compact accelerators are developed for wider applications, such as in medicine. And the impact on the way particle physics is done in this century will no doubt be profound.

Robert Bingham is at the Rutherford Appleton Laboratory, Chilton, Didcot OX11 0QX, UK and the Physics Department, University of Strathclyde, Glasgow G4 0NG, UK.
e-mail: r.bingham@rl.ac.uk

1. Blue, B. E. *et al.* *Phys. Rev. Lett.* **90**, 214801 (2003).
2. Geraci, A. & Whittum, D. *Phys. Plasmas* **7**, 3431–3440 (2000).
3. Lee, S., Katsouleas, T., Hemker, R. G., Dodd, E. S. & Mori, W. B. *Phys. Rev. E* **64**, 045501(R) (2001).
4. Malka, V. *et al.* *Science* **298**, 1596–1600 (2002).
5. Lee, S. *et al.* *Phys. Rev. Spec. Topics Accel. Beams* **5**, 011001 (2002).
6. Ng, J. S. T. *et al.* *Phys. Rev. Lett.* **87**, 244801 (2001).

biochemical processes that determine lifespan? The findings reported by Murphy *et al.* on page 277 of this issue² represent an important step towards the answer. Using DNA microarray technology, these authors have identified a large set of genes whose activity is linked to lifespan, and they provide evidence that many of these genes act in concert to control longevity and ageing.

A prime example of an interesting gerontogene that had, until now, provided little insight into the mechanics of ageing is *daf-16*. This gene encodes a transcription factor, DAF-16—a protein that modifies the activity of other genes—which is a powerful regulator of *C. elegans* lifespan^{3,4}. DAF-16 is switched off by a hormonal signalling pathway akin to that activated by the mammalian insulin and insulin-like growth factor 1 (IGF-1) proteins⁵. Reduced activity of this pathway can greatly increase adult life span not only in *C. elegans*⁶, but also in fruit-flies⁷ and mice⁸. Even ten years ago, it was clear that to understand ageing thoroughly we would need to identify the genes regulated by DAF-16 (ref. 6). Yet screening for mutants failed to identify any of these 'downstream' genes.

Fortunately, a new technology has emerged that is well suited to addressing this question: DNA microarrays. Using microarrays, researchers can take a snapshot of the activity of most of the genes in the fully sequenced *C. elegans* genome. By comparing the snapshots from normal animals to similar snapshots from animals in which a particular gene is mutated, it is possible to work out which other genes in the mutants are activated or repressed as a result of that mutation. Murphy *et al.* used this technology to identify two classes of genes that are activated or repressed by DAF-16. Class 1 genes are switched on by DAF-16 and are associated with increased lifespan, whereas class 2 genes are repressed by DAF-16, and are associated with reduced lifespan. Using these analytical approaches, the authors found 189 class 1 and 122 class 2 genes.

But do any of these genes actually have a role in determining lifespan? To test this, Murphy *et al.* analysed the genes' functions, using an elegant methodology called RNA interference. If *C. elegans* are fed bacteria that produce pieces of RNA matching a given worm gene, the activity of that gene is reduced, or silenced. The expectation was that in some cases, silencing the genes in class 1 would reduce lifespan, whereas silencing class 2 genes would increase lifespan. As it turned out, a surprisingly high proportion of genes behaved as predicted.

This is an important finding. The previous failure to identify genes that act downstream of DAF-16 had raised the gloomy prospect that this protein might regulate numerous genes that act in concert, such that inactivation of any single gene would

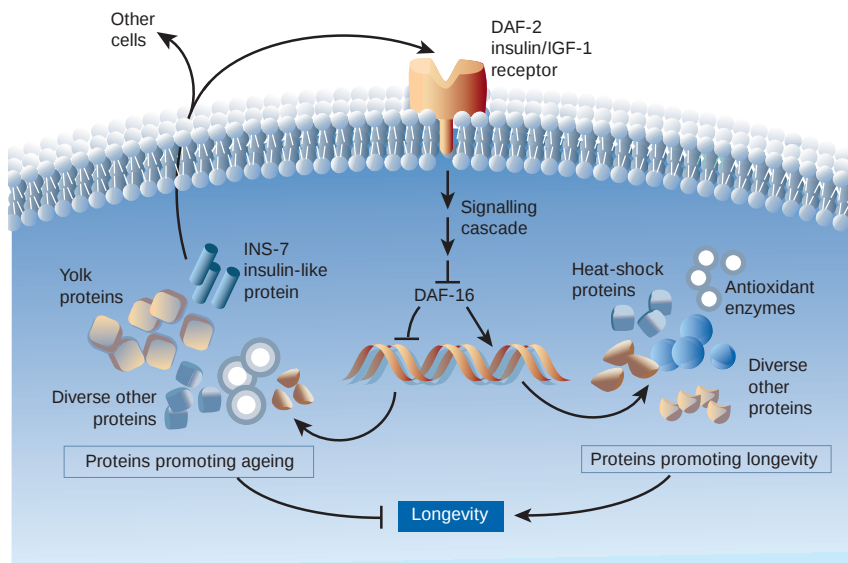


Figure 1 Ageing versus long life: the molecules involved. Murphy *et al.*² have found that the transcription factor DAF-16 controls the expression of a battery of genes, many of which have small effects on lifespan—promoting either ageing or longevity—in *Caenorhabditis elegans*. Consistent with earlier studies, the pro-longevity genes include some that encode antioxidant enzymes and others encoding heat-shock proteins, which can restore misfolded proteins to their active conformations. Genes that promote ageing include some that encode yolk proteins, consistent with a link between ageing and reproduction. Another pro-ageing protein is the insulin-like INS-7, which, by binding to the insulin/IGF-1 receptor (DAF-2), may repress DAF-16 on the same and other cells. This suggests the presence of a positive feedback loop that regulates DAF-2 activity. There are also many other proteins whose mechanistic links to ageing are as unclear as they are intriguing. Arrows indicate activation; T-bars indicate inhibition.

have no perceptible effect. Instead, it appears that the increased longevity resulting from DAF-16 activation is due to the additive effects of many genes, which individually exert a small effect on lifespan. This is a satisfying discovery, because if individual downstream gerontogenes had little effect on ageing, it would be difficult to see how longevity could evolve.

So does the identity of DAF-16-regulated genes explain the biochemistry of ageing, or is it just another big list of gerontogenes? The list certainly gives a few inklings. One influential theory of the biochemistry of ageing is the oxidative-damage theory: over time, oxidative by-products of metabolism cause molecular damage, which accumulates and eventually causes ageing and death⁹. Consistent with another recent microarray study involving *daf-16* (ref. 10), Murphy *et al.*² found that some of the genes that are upregulated by DAF-16 do encode enzymes that protect against or repair oxidative damage (Fig. 1).

But numerous other biochemical processes are implicated; DAF-16 also regulates genes involved in antimicrobial responses, genes for several protein-digesting enzymes, metabolic genes, and many genes of unknown function. Finally, several signalling molecules seem to influence longevity, including an insulin-like protein encoded by *ins-7*—one of the bewildering multiplicity

of 39 such genes in the *C. elegans* genome. Potentially, the most important proteins identified are among the many whose mechanistic link to ageing is mysterious. It is tantalizing to think that among these proteins could be some that determine the differences in longevity between animal species, and control human ageing.

Seeking insight into the biochemistry of ageing, Murphy *et al.* panned their mine of microarray data for genes that fit existing expectations—an approach sometimes referred to as 'fishing'. But the drawback with fishing is that it is all too tempting to fit your data to the hypothesis, rather than the other way around. What is needed now is an unbiased and statistically rigorous analysis of which functional classes of genes are controlling lifespan. Such an approach would be better able to identify processes not previously known to be involved in ageing. The raw microarray data generated by Murphy *et al.* are publicly available, so other researchers should be able to analyse the data further and fully realize the value of this information.

Understanding the function of the impressive list of genes identified here², and their role in ageing, must be the next major task—one likely to take years of painstaking analysis. This will mean much hard work at the molecular-biology coalface, after the heady interlude of microarray analysis. ■

David Gems and Joshua J. McElwee are in the Department of Biology, University College London, Gower Street, London WC1E 6BT, UK.
email: david.gems@ucl.ac.uk

1. Partridge, L. & Gems, D. *Nature Rev. Genet.* **3**, 165–175 (2002).
2. Murphy, C. T. et al. *Nature* **424**, 277–284 (2003).
3. Ogg, S. et al. *Nature* **389**, 994–999 (1997).
4. Lin, K., Dorman, J. B., Rodan, A. & Kenyon, C. *Science* **278**, 1319–1322 (1997).

5. Kimura, K. D., Tissenbaum, H. A., Liu, Y. & Ruvkun, G. *Science* **277**, 942–946 (1997).
6. Kenyon, C., Chang, J., Gensch, E., Rudener, A. & Tabtiang, R. *Nature* **366**, 461–464 (1993).
7. Tatar, M. et al. *Science* **292**, 107–110 (2001).
8. Holzenberger, M. et al. *Nature* **421**, 182–187 (2003).
9. Sohal, R. S. & Weindruch, R. *Science* **273**, 59–63 (1996).
10. McElwee, J., Bubb, K. & Thomas, J. H. *Aging Cell* **2**, 111–121 (2003).

Global change

The past and future of El Niño

Sandy Tudhope and Mat Collins

A new study of past variations in El Niño behaviour provides a much improved record from pre-instrumental times. It will be a valuable resource for testing the models used in climate prediction.

Droughts, floods, forest fires and changed patterns of storms and ocean conditions — these are some of the hallmarks of the El Niño Southern Oscillation (ENSO) climate cycle. On page 271 of this issue, Cobb *et al.*¹ describe how their analysis of fossil corals provides new evidence of how the ENSO cycle can strengthen and weaken without an obvious external driving force. Not least, these data provide helpful indications as to how ENSO might change in the future.

El Niño events occur every three to seven years, and arise naturally through the strong inter-action between the ocean and atmosphere in the tropical Pacific². The effects are felt worldwide because the tropical Pacific is a powerful source of heat for driving atmospheric circulation. Even small changes in the sea-surface temperature in this region have major repercussions for global climate. Dry conditions in the western Pacific, warmer-than-average conditions in the northern United States and wetter-than-average conditions to the south, and suppressed hurricane development in the Atlantic, are all indicators of an El Niño event. Consequently, when the ENSO cycle is particularly strong, as occurred in 1982–83 and again in 1997–98, the changed weather patterns have profound social, economic and ecological consequences³.

Although much of the basic physics of the ENSO cycle is now reasonably well known, some aspects remain poorly understood. What, for instance, causes one El Niño event to be different from another? One obstacle to progress is the paucity of instrumental data before the mid-twentieth century. From the data that do exist, it is clear that ENSO varies considerably in strength and frequency, but the instrumental records are too short to tell with confidence if this variability is internal or driven by changes in background climate⁴.

Cobb *et al.*¹ use the skeletons of annually banded reef-building corals as natural

'proxy' archives of tropical climate, and thereby extend the record of ENSO further into the past. Their approach involves analysing the oxygen isotopic composition of the coral skeletons as a measure of the combined effects of changes in sea-surface temperature (through a temperature-dependent fractionation between water and calcium carbonate), and seawater isotopic composition (which is strongly influenced by rainfall and evaporation). At their study site, Palmyra Island in the central tropical Pacific, warm temperatures and increased rainfall during the El Niño phase of the Southern Oscillation combine to drive the oxygen isotopic composition of the coral skeletons towards lighter (more negative) values. The excellent match between an isotope record from a living coral and the instrumental records of recent ENSO activity provides proof of concept.

However, living corals yield only about 100 years of record, so Cobb *et al.* also analysed cores collected from ancient corals, washed onto the top of the reef by storms. Using an approach borrowed from tree-ring reconstructions of climate, they combine high-precision dating with 'wobble matching' of the monthly-resolution isotope records to create continuous coral chronologies. The final record is not complete: they have reconstructed several periods, representing a total of some 430 years since about AD 930, but the method has yielded much more robust, and extended, ENSO proxy records than have been produced before for this period.

The key finding is that ENSO appears to have varied considerably in strength over the past millennium, with changes occurring rapidly, on timescales of decades or so. Compared with earlier periods, the events in the twentieth century appear to have been relatively strong, but not exceptionally so. The inferred changes in ENSO strength do not appear to be correlated with documented changes in Northern Hemisphere regional climate — for instance, the Little Ice Age



100 YEARS AGO

Rutherford and Soddy pointed out that the almost invariable presence of helium in minerals containing uranium indicated that that gas might be one of the ultimate products of the disintegration of the radio-elements. Rutherford, moreover, determined the mass of the projected particle which constitutes the "α-ray" of radium to be approximately twice as great as that of the hydrogen atom, an observation that points in the same direction... We have been engaged for some months in examining the spectrum of the "radio-active emanation" from radium... We have found that after removing hydrogen and oxygen from the gases evolved from 20 mgrs. of radium bromide, the spectrum showed the presence of carbon dioxide. On freezing out the carbon dioxide, and with it, a large proportion of the radium "emanation," the residue gave unmistakably the D₃ line of helium.

ALSO...

It has been stated that the radium rays have been successfully applied in the treatment of a case of cancer by Prof. Gussenbauer, of Vienna. The tumour completely disappeared as a result of the application, radium bromide being made use of as a source of the rays. The early publication of these details in the public Press before there has been time to test the method effectually is much to be deprecated.

From *Nature* 16 July 1903.

50 YEARS AGO

Personality Development. By J. S. Slotkin. The author, who is a social anthropologist, begins by saying that he has tried to work out a systematic theory of personality development, from the hypotheses and evidence of various relevant sciences. The material is, however, systematic only in so far as it has been distributed under four major headings: inheritance, socialization, culturization and individualization. Within each of the fields the approach is mainly descriptive and anecdotal, and is adorned by a wealth of quotations from ancient and modern writers. It may well be that in the meantime this kind of approach is repaying. Certainly it does not suffer from the aridity of some recent statistical work in the same field. But perhaps we can only be scientific about people if we are content to be dull.

From *Nature* 18 July 1953.

(seventeenth to nineteenth centuries) or the Medieval Warm Period (eleventh to fourteenth centuries). Nor do they seem to tie in with reconstructions of volcanic and solar behaviour that might drive climate change. The authors conclude, quite reasonably, that much of the variability seen in ENSO strength over the past millennium was probably not driven by external factors.

So, where does this leave us in attempting to predict future ENSO activity? Cobb *et al.* show that, as ENSO varies significantly on its own — that is, even without greenhouse warming — we might expect changes beyond those experienced in the twentieth century. Looking to other periods in the past there is increasing evidence that the ENSO cycle may have been weak, or even absent, between about 14,000 and 5,000 years ago^{5,6}. The most likely explanations involve the sensitivity of ENSO to changes in the length and timing of the seasons resulting from the precession cycle in the Earth's orbit⁷. ENSO may also have been generally weaker during the cooler conditions of the last glacial period (100,000–20,000 years ago)⁶. These data hint that the ENSO system is sensitive to background climate, and that it may well respond to future greenhouse warming. But it is not easy to tell what this response will be — there is no past equivalent of expected twenty-first-century climate.

Cobb *et al.* looked at ENSO during a period of relatively little global climate change. The future is almost certainly going to be radically different, with predictions of a 1.5–6.0 °C rise in global mean temperature within the next 100 years⁸ — potentially almost as much change as that between the last glacial period and the present interglacial. The only realistic hope for predicting the response of ENSO to this warming lies in the use of coupled ocean–atmosphere climate models. Some of the best of these models now generate a realistic ENSO cycle but produce a wide range of predicted outcomes for ENSO in a warmer world, from a significant strengthening of the cycle, to no effect or even a weakening^{9,10}. Equally important in terms of potential social and economic consequences are changes in ‘average’ conditions, for example towards a more El Niño-like background state with associated reduction in rainfall over much of southeast Asia and South America¹¹.

The use of proxy data about past conditions has been limited in evaluating global ocean–atmosphere climate models¹², partly because modellers find it difficult to deal with isolated, short and imprecisely dated records like those from fossil corals. But such tests are crucial. The study of Cobb *et al.* is a step towards the more complete multi-proxy¹³ archives of climate that will be needed.

Sandy Tudhope is at the School of GeoSciences, University of Edinburgh, West Mains Road, Edinburgh EH9 3JW, UK.

e-mail: sandy.tudhope@ed.ac.uk

Mat Collins is at the Centre for Global Atmospheric Research, Department of Meteorology, University of Reading, Reading RG6 6BB, UK.

e-mail: matcollins@met.rdg.ac.uk

1. Cobb, K. M., Charles, C. D., Cheng, H. & Edwards, R. L. *Nature* **424**, 271–276 (2003).
2. Philander, S. G. H. *El Niño, La Niña and the Southern Oscillation* (Academic, San Diego, 1990).
3. Glantz, M. H. *Currents of Change: El Niño's and La Niña's Impact on Climate and Society* (Cambridge Univ. Press, 2001).

4. Folland, C. K. *et al.* *IPCC Third Assessment Report Ch. 1* (Cambridge Univ. Press, 2001).
5. Rodbell, D. T. *et al.* *Science* **283**, 516–520 (1999).
6. Tudhope, A. W. *et al.* *Science* **291**, 1511–1517 (2001).
7. Clement, A. C. *et al.* *Paleoceanography* **15**, 731–737 (2000).
8. Cubash, U. & Meehl, G. A. *IPCC Third Assessment Report Ch. 9* (Cambridge Univ. Press, 2001).
9. Timmermann, A. *et al.* *Nature* **398**, 694–696 (1999).
10. Collins, M. *Geophys. Res. Lett.* **27**, 3509–3513 (2000).
11. Cox, P. M. *et al.* *Nature* **408**, 184–187 (2000).
12. Collins, M. *et al.* *J. Clim.* **15**, 1497–1515 (2002).
13. Mann, M. E. *et al.* *Nature* **392**, 779–787 (1998).

Quantum physics

Uncertain future

Miles Blencowe

The uncertainty principle limits the accuracy of measurement at the quantum level. A device sensitive to subatomic-scale displacement has come close to revealing that principle in action in the macroscopic world.

In 1927, Werner Heisenberg introduced his famous quantum principle¹, which states that the uncertainties in the position and the velocity of a particle are inversely proportional to each other: a particle's position or its velocity can be known precisely, but not both at once. This principle is one of the cornerstones of quantum mechanics, and is traditionally relevant to the domain of subatomic particles. But what about more familiar macroscopic objects, comprising many atoms, that we think of as possessing simultaneously well-defined positions and velocities of their centre-of-mass? If we could be sufficiently

precise in our measurements on such objects, would we encounter the quantum uncertainty principle at work?

On page 291 of this issue, Knobel and Cleland² address this question. They describe an exquisitely engineered device — a vibrating crystal beam, only a thousandth of a millimetre long, and an extremely sensitive motion detector — that is capable of detecting displacements as small as about one-thousandth of a nanometre, or one-hundredth of the size of a single atom. The beam may seem tiny by everyday standards, but its mass is equivalent to that of about ten billion atoms. A demonstration of the

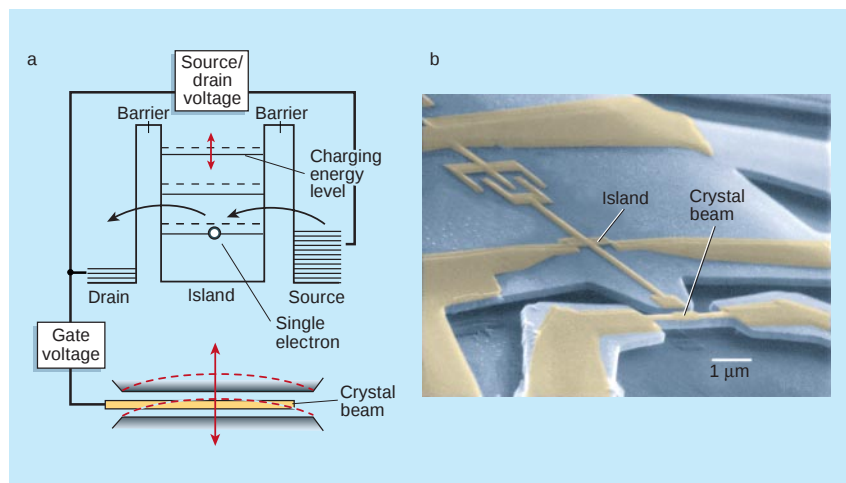


Figure 1 Closer to the limit. Knobel and Cleland² have built a device, based on a single-electron transistor, to investigate the effect of Heisenberg's uncertainty principle on a macroscopic system. a, Electrons tunnel from the source to the island and then to the drain, constituting a flow of current. The voltage between source and drain is only strong enough to charge the island by one electron at a time (as indicated by the position of the source level relative to the island charging levels). If the gate voltage is varied, this causes the island charging levels to shift, changing the magnitude of the tunnelling current. Alternatively, fixing the gate voltage and allowing the crystal beam to bend also shifts the levels and modulates the tunnelling current. b, The device, shown here in an electron micrograph, is only micrometres in size, and although the quantum zero-point detection limit has not yet been reached, the experiment is the best demonstration so far of sensitivity to subatomic-scale displacements of a nanoscale beam.

uncertainty principle for such a system would be a striking manifestation of quantum mechanics well outside its traditional, microscopic domain.

If you clamp one end of a wooden ruler to the edge of a table and then pluck the other, free end, it vibrates with decaying amplitude and eventually returns to apparent rest. But if you were to look at the free end of the ruler under a sufficiently powerful microscope, it would not be at rest at all, but jiggling up and down in a random fashion. This motion is a consequence of the air molecules striking the ruler, as well as of its countless, fluctuating internal defects, and is an example of thermal brownian motion.

There are other, quantum fluctuations in the ruler, though, that are completely masked by this classical thermal motion. These quantum 'zero-point' fluctuations have much smaller amplitude and arise from the necessary uncertainty in position and velocity stated in Heisenberg's principle. The situation is analogous to the hydrogen atom, which is stable because the attractive electrostatic force that would like to pull the electron into a tighter volume around the proton is balanced by the repulsive effect of the electron's fluctuating velocity. Similarly, for a macroscopic object such as a crystal beam or a ruler, the elastic restoring force on the bent beam balances the repulsive effect of its fluctuating centre-of-mass velocity.

Because the magnitudes of the zero-point fluctuations in position and velocity are so small, they can only be detected if the structure is cooled down to very low temperatures. As the temperature is lowered, the amplitude of thermal motion decreases. Eventually, there will be no thermal motion, only pure, temperature-independent zero-point fluctuations. At a temperature of about a hundredth of a kelvin (close to the temperature Knobel and Cleland achieve in their experiment²), zero-point fluctuations should dominate in a structure with a mechanical vibration frequency of about one billion cycles per second (1 gigahertz, or GHz). It is well known that the pitch of a plucked violin or guitar string increases with decreasing length, so it's easy to appreciate why such a structure must be extremely small. A tiny, 1-GHz, crystal beam was demonstrated recently³; Knobel and Cleland's beam vibrates at a frequency only about ten times lower, at the upper end of the FM radio band.

Knobel and Cleland estimate that their crystal beam has a zero-point displacement uncertainty of about 10^{-5} nanometres. To detect such small displacements, they use a device called a single-electron transistor⁴, which consists essentially of a small island of metal, comparable in size to the crystal beam, separated by electrically insulating gaps from two wire leads that are connected to a voltage source (Fig. 1). The gaps are sufficiently

narrow that electrons can tunnel across them, from one lead (the source) to the other (the drain), through the island. The voltage difference between the leads can be set so that only one electron at a time can tunnel onto the island: an electron on the island must tunnel off into the drain before another tunnels from the source to the island.

This single-electron current is extremely sensitive to charge fluctuations in the vicinity of the island, a property that Knobel and Cleland exploit to turn this single-electron transistor into a displacement detector. The metal island and the crystal beam in their device are separated by a vacuum gap of about 250 nanometres. The beam has a thin metal coating and a separate voltage applied to it (the gate voltage), coupling the beam electrostatically to the island. As the beam centre-of-mass position and hence the gap between the beam and island fluctuate, the charges on the island redistribute. This in turn causes a fluctuation in the tunnelling current that can be measured and related to the displacement of the beam.

Although Knobel and Cleland's device represents a milestone in fast and ultrasensitive displacement detection, it still doesn't quite take us to the quantum zero-point detection limit. To do so, the device's sensitivity to displacements must be improved by about a factor of a hundred, and the mechanical vibration frequency of the beam increased by about a factor of ten. Nevertheless, Knobel and Cleland's achievement throws out an exciting challenge — to close the gap of these remaining few orders of magnitude, and demonstrate Heisenberg's uncertainty principle at work in the macroscopic domain. ■

Miles Blencowe is in the Department of Physics and Astronomy, 6127 Wilder Laboratory, Dartmouth College, Hanover, New Hampshire 03755, USA. e-mail: miles.p.blencowe@dartmouth.edu

1. Heisenberg, W. Z. *Phys.* **43**, 172–198 (1927); English translation in *Quantum Theory and Measurement* (eds Wheeler, J. A. & Zurek, W. H.) 62–84 (Princeton Univ. Press, 1983).
2. Knobel, R. G. & Cleland, A. N. *Nature* **424**, 291–293 (2003).
3. Huang, X. M. H. *et al.* *Nature* **421**, 496 (2003).
4. Devoret, M. H. & Schoelkopf, R. J. *Nature* **406**, 1039–1046 (2000).

Evolutionary biology

Body plans and simple brains

Thurston Lacalli

Genes expressed in the vertebrate brain and spinal cord show up in the surface nerve net of a closely related group of invertebrates. Could this mean that brains started out on the body surface?

As a vertebrate embryo grows, the development of its brain and spinal cord is controlled by complex and precisely regulated patterns of gene activity. Writing in *Cell*¹, Lowe *et al.* now report similar patterns from the embryos of hemichordates, which are invertebrates but are related to vertebrates and may resemble their ancient ancestors. This observation has implications for our understanding of how an internal brain and spinal cord first arose in vertebrates, and of other changes to their body plan that are controlled by the same genes.

The genes in question are responsible for patterning the body along its antero-posterior axis (that is, from front to back), notably Hox genes and the like. These genes are highly conserved in evolution, with similar expression patterns in animals as anatomically different as insects and vertebrates. Insects and vertebrates are advanced members, respectively, of the two major divisions of animals, protostomes and deuterostomes (Fig. 1, overleaf). Evolutionary biologists have long wanted to know what the last common ancestor of these two groups was really like. Was it segmented, as are both insects and vertebrates? Did it have a condensed, internal nervous system,

as both do? Or was it much simpler, so that the similarities between its various descendants are due to evolutionary convergence? These are key questions relating to the nature of the genetic programme that underlies embryonic development and how it constrains evolution.

The enigmatic last common ancestor is now long extinct. But among modern deuterostomes there are surviving groups that are close to it — notably echinoderms, hemichordates and tunicates, the last being the most 'basal' chordates (the grouping to which vertebrates belong; Fig. 1). Enteropneust hemichordates — also known as acorn worms because of their bulbous proboscis — are of special interest. That's because, alone among the basal deuterostomes, they have a simple bilateral body plan (worm-like with a proper front and back end) and a comparatively active mode of life as burrowers in marine sediments. They thus provide a plausible model for the proximate ancestor of chordates. Among the other candidates, echinoderms and tunicates show varying tendencies to adopt a sedentary habit and a reduced or substantially modified (for instance, radial) body plan, which seriously complicates interpretation.

Hemichordates are obscure animals, but

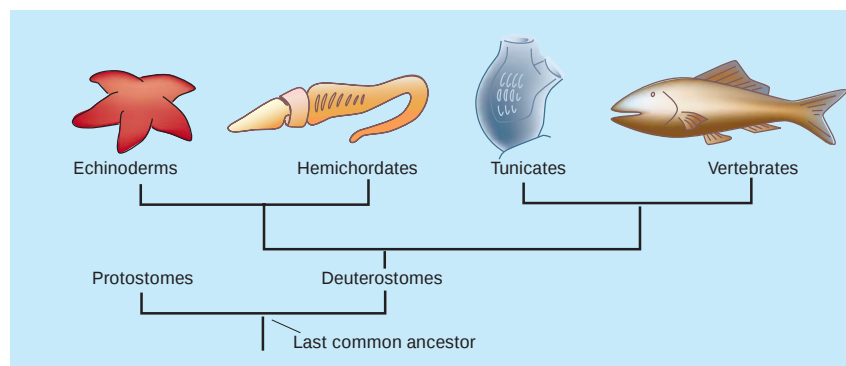


Figure 1 Protostomes and deuterostomes. Evolutionary biologists are especially interested in whether any of the surviving 'basal' deuterostomes (echinoderms, hemichordates and tunicates) resemble the last common ancestor of these two major divisions, or whether (like vertebrates) they have been much modified by subsequent evolution. *Saccoglossus*, the creature whose development is studied by Lowe *et al.*¹, is a hemichordate.

they have been the subjects of an increasing number of molecular studies as their phylogenetic importance has become more widely recognized. Lowe *et al.*¹ chose *Saccoglossus*, a species that develops directly into the adult form without going through a larval phase. Their study is particularly ambitious. They describe the expression patterns of 22 genes, and then discuss the implications for competing ideas of chordate origins. These results also figure prominently in a new survey of neural evolution from its most primitive beginnings².

The most surprising aspect of Lowe and colleagues' results is that axial-patterning genes in *Saccoglossus* are expressed throughout the surface tissue, the ectoderm, in a series of domains that encircle the whole body (Fig. 2). This is more like the situation in insects than in vertebrates, where ectodermal expression is confined to the domain fated to form the internal nerve cord. Hemichordates lack a direct equivalent of the nerve cord, however, and instead have a diffuse nerve net that lies within the epithelium and is spread over the body surface. So the rule would seem to be that patterning genes are expressed in the nerve-forming tissue domain regardless of the location and extent of that domain.

The second surprise is that the patterns of gene expression are nearly identical in hemichordates and vertebrates (and very similar to those in insects), even though vertebrates and insects have nervous systems that are vastly more complex anatomically than those of hemichordates. Thus, the genetic map for axial patterning is complex and apparently unchanged as far back as the earliest bilateral animals, irrespective of morphology. Lowe *et al.* favour the idea that hemichordates preserve the ancestral condition of simple morphology but a complex genetic map. If so, internal nerve cords would have had to evolve independently in the ancestors of both insects and vertebrates.

The converse is also possible — that an internal ancestral nervous system reverted to the surface in hemichordates as they took to burrowing. Where the whole body is in contact with the surrounding substratum, a diffuse nerve net of maximal extent might have the advantage. One therefore wants to know more about the functional organization of the hemichordate system to see, for example, if it is truly primitive in organization or only appears so. Much of the basic neurobiology, however, still remains to be done. There is also a link here to vertebrate brains, in that many basal brainstem centres in vertebrates have a net- or plexus-like organization³. This

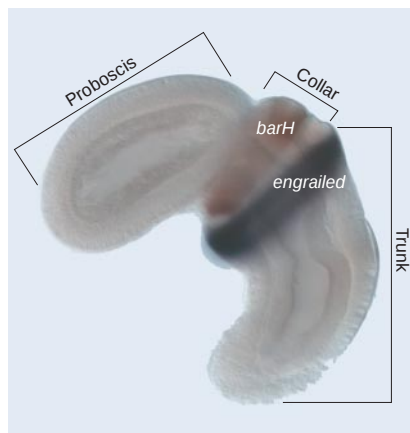


Figure 2 *Saccoglossus* embryo showing expression patterns for two of the 22 genes examined by Lowe *et al.*¹. In vertebrates, the *engrailed* gene marks the midbrain-hindbrain junction. Here it forms a band at the front of the trunk region around the circumference of the body (hemichordates lack a brain as such). The *barH* gene is expressed further forward in vertebrate brain, through the caudal forebrain and midbrain. Here it is also further forward, one of numerous examples¹ where gene-expression patterns in vertebrate brain and spinal cord map almost unchanged to the body surface of the embryonic hemichordate. In life, the embryo is about 700 μm long.

could conceivably reflect an origin from a superficial nerve net performing a similar function to that in hemichordates — which, again, makes the latter uniquely informative from a vertebrate perspective.

In the near term, perhaps the most valuable contribution of the new results¹ is to place the comparative study of how the body and brain are regionally subdivided in deuterostomes on a much firmer footing. For example, the combination of genes that specify the junction of midbrain and hindbrain in vertebrates lines up with the first of the gill slits, structures that are found in both vertebrates and *Saccoglossus*. This implies that wherever gill slits occur near the back of the body, as they do in adult tunicates and hemichordate larvae in species that have them, it is an indication that everything anterior to that point is essentially head-like in character. Thus the tunicate body, in the adult, can be considered a sedentary 'head', despite being packed with visceral organs and gonads that would originally have resided elsewhere. Likewise, hemichordate larvae are apparently, in large part, small swimming heads. This tends to discredit hypotheses, such as that of Garstang, that derive the whole of the chordate body from such larvae⁴. Instead, one would now predict that only very anterior head structures could arise in this fashion, and there is some independent molecular evidence that this may be the case⁵.

A logical next step from anteroposterior axes is to dorsoventral ones. Lowe and colleagues are working on the genes involved, in *Saccoglossus*, in the hope of resolving the controversy over dorsoventral inversion. The issue here is that insects, with their ventral nerve cord, and vertebrates with their dorsal one, appear to be inverted with respect to one another. There is now convincing evidence that this goes beyond anatomy to the molecular level⁶. When and how the inversion occurred in evolution, with the concomitant rearrangement required of other structures such as the mouth, remains very much a puzzle. To date, hemichordates have been left somewhat outside this debate owing to the rather ambiguous evidence as to which side of their body is actually dorsal⁷. The molecular work now in progress on *Saccoglossus* should remove this ambiguity, and help pave the way for a resolution of the overall issue.

Thurston Lacalli is at the Department of Biology, University of Victoria, Victoria, British Columbia V8W 3N5, Canada.
e-mail: lacalli@uvic.ca

1. Lowe, C. J. *et al.* *Cell* **113**, 853–865 (2003).
2. Holland, N. D. *Nature Rev. Neurosci.* doi:10.1038/nrn1175 (2003).
3. Nieuwenhuys, R. *Prog. Brain Res.* **125**, 49–126 (2000).
4. Gee, H. *Before the Backbone* (Chapman & Hall, London, 1996).
5. Tagawa, K. *et al.* *Evol. Dev.* **3**, 443–454 (2001).
6. Gerhart, J. *Proc. Natl. Acad. Sci. USA* **97**, 1113–1122 (2000).
7. Ruppert, E. E. *Invert. Biol.* **118**, 202–212 (1999).

Microscopy

High-speed movies in miniature

Appl. Phys. Lett. **83**, 6–8 (2003)

Microscopic movies have been given a higher frame speed by Andrew D. L. Humphris and colleagues. This means that optical microscopy can be used to follow faster processes in, for example, living cells.

The scanning near-field optical microscope (SNOM) developed by Humphris *et al.* can image an area of $20\ \mu\text{m}^2$ in less than 10 ms, at a speed of more than 100 frames per second. This is more than a thousand times faster than commercial SNOMs, and ten times faster than any scanning probe microscope reported previously.

SNOMs, which can generate optical images with a resolution of less than a micrometre (that is, below the normal limit imposed by diffraction effects), are widely used for *in vivo* imaging in cell biology. The new improvement in speed hinges on two issues: greater imaging bandwidth, which increases the signal intensity and thus decreases the data acquisition time, and a new scanning mechanism that exploits a large-amplitude mechanical resonance of the instrument to move the probe tip (an optical fibre) back and forth. Normally, this kind of resonant oscillation is regarded as a nuisance that must be suppressed. **Philip Ball**

Cell biology

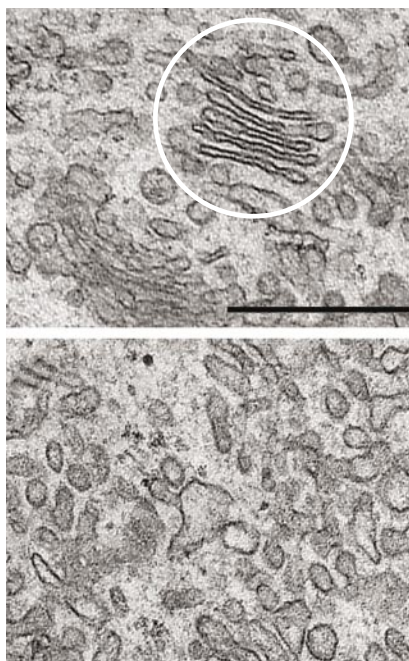
Golgi glue

EMBO J. **22**, 3279–3290 (2003)

A stack of flattened membranes — not unlike a pile of pitta bread — makes up the Golgi apparatus in a cell. Many newly made proteins travel through this system to their final cellular destinations. En route the proteins are modified by enzymes, which are arranged across the stacks in the order in which they function. So it's important that the Golgi remains well organized. Yanzhuang Wang *et al.* now find that the protein GRASP65 may be the glue that keeps the stacks together.

One school of thought holds that the Golgi apparatus breaks up during cell division to allow its membranes to be distributed to both daughter cells. Subsequently, each cell reassembles a new Golgi from the parts. In support of this, Wang *et al.* show, by recreating the phases of the cell cycle in a test tube, that GRASP65 is phosphorylated — and presumably out of action — during cell division, but is necessary for proper stacking of the Golgi afterwards (see picture).

The authors also find that GRASP65 can form dimers, and that beads coated with GRASP65 dimers aggregate. The beads



Cellular pitta bread: top, normal Golgi (circled); bottom, what happens when GRASP65 is blocked in cell division. Scale bar, $0.5\ \mu\text{m}$.

separate after enzymes that phosphorylate GRASP65 during cell division are added, but reaggregate upon its dephosphorylation. With a stickiness that seems to alter with the cell cycle, it looks like GRASP65 has what it takes to be Golgi glue. **Marie-Thérèse Heemels**

Stem cells

Turning back the clock

Curr. Biol. **13**, 1206–1213 (2003)

Researchers have been trying for some time to decipher the molecules and mechanisms that can turn adult cells into youthful stem cells. Now John B. Gurdon and colleagues report that immature frog eggs can rejuvenate human and mouse nuclei alike.

The authors injected the oocytes of *Xenopus* frogs with nuclei taken from adult mouse or human white blood cells. Two days later, mouse or human *Oct-4* messenger RNA — the definitive stem-cell marker — had appeared. This indicates that the adult DNA had been reprogrammed in a species-specific way to become stem-cell-like. The effect was enhanced when the genetic material was injected directly into the oocyte nucleus, leading the team to speculate that a nuclear component may be responsible for the eggs' revitalizing abilities.

Gurdon and colleagues hope to analyse and isolate the molecules responsible so that, in future, adult cells taken from patients can be reprogrammed directly. This would allow the production of a limitless supply of donor-matched stem cells with which to repair and replace damaged and diseased tissue. **Helen R. Pilcher**

Acoustics

New bells ring sweeter

J. Acoust. Soc. Am. **114**, 505–511 (2003)

Bells are rarely used in Western music, at least outside of orchestral 'character roles', because they typically produce overtones that may sound dissonant when combined with other instruments. In particular, European bells tend to generate a strong minor-third tone, which could wreak havoc within a complex chord. Neil McLachlan *et al.* have used shape-modelling calculations to devise bells with overtones that are 'pure' harmonics, eliminating these problematic resonances.

Attempts to make harmonically tuned bells date back to the seventeenth century, but have had mixed success. The current researchers use finite-element modelling to deduce wall-thickness profiles of conical bells that produce up to six harmonics above the fundamental (or 'hum') tone. Their designs have been implemented by Australian Bell Ltd, which has made bronze bells cast and lathed to fit the calculated profiles. These harmonic bells have already been used in musical performances with the Melbourne Symphony Orchestra, and McLachlan and colleagues anticipate that their designs will "expand the musical applications of bells". **Philip Ball**

Diabetes

Molecular suspect fingered

Dev. Cell **5**, 73–83 (2003)

Type 2 diabetes mellitus will become one of the world's largest public-health problems, but we know little about its underlying biochemistry. A new molecular culprit has now been implicated in the disease.

The symptoms of type 2 diabetes are caused by high levels of blood sugar, which arise because β -cells in the pancreas become dysfunctional and stop releasing insulin normally. But what are the molecular signals that cause β -cells to malfunction?

Bruce M. Spiegelman and colleagues provide a clue. They have found that increased levels of a single protein — PGC-1 α — are sufficient to induce diabetes-like effects in pancreatic cells. Moreover, levels of this protein are higher than normal in various animal models of the disease. PGC-1 α appears to directly regulate the activity of several different metabolic genes, which in turn affect the ability of β -cells to respond to glucose and release insulin as necessary.

The same authors previously found that PGC-1 α also activates genes involved in glucose production in the liver. So it seems that PGC-1 α is a powerful regulator of both glucose production and insulin release. **Clare Thomas**

Birds sing at a higher pitch in urban noise

Great tits hit the high notes to ensure that their mating calls are heard above the city's din.

The ongoing spread of urban areas, highways and airports throughout the world makes anthropogenic noise almost omnipresent. We have found that urban great tits (*Parus major*) at noisy locations sing with a higher minimum frequency, thereby preventing their songs from being masked to some extent by the predominantly low-frequency noise. They have presumably learned selectively from a restricted range of their repertoire — a behavioural plasticity that may be critical for breeding success in a noisy world.

Cars, planes and all sorts of machinery create a new selection pressure on wildlife species that use acoustic signals to achieve reproductive success. This might create two groups of species: one that can adapt their signals to the competing noise, and another that cannot. Although there is a decline in species density and diversity associated with sprawling cities and highways^{1–4}, there is no evidence yet for a direct role of sound pollution^{5,6}, nor is there much insight into how successful urban species cope under noisy circumstances^{7,8}.

We investigated an urban population of great tits in the Dutch city of Leiden. Noise-amplitude measurements, taken with a sound-pressure meter, varied markedly between territories. Mean amplitude levels per territory ranged from 42 to 63 decibels, from very quiet in residential areas to extremely noisy near a highway or a busy crossing. We used a highly directional microphone for song recordings and an omnidirectional microphone for independent noise recordings at a height of 5 m. The spectral composition of ambient noise was generally characterized by loud, low-frequency sounds.

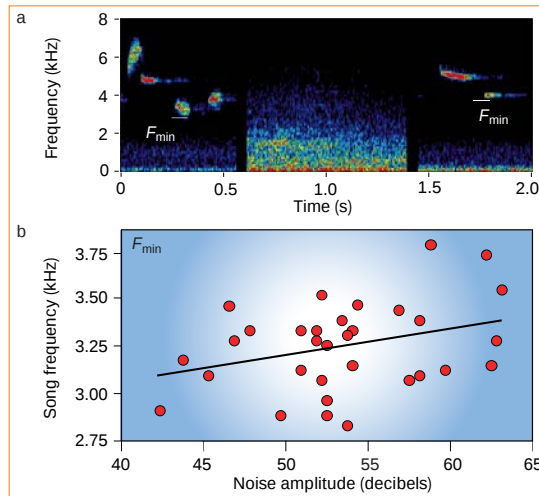
We compared noise amplitude with the spectral distribution of sound energy within the range of the minimum frequency of great-tit song and found that in noisy territories there is a greater proportion of sound energy in the lower half of this range than in quiet territories (Pearson's $r = 0.78$, $P < 0.001$).

We measured the acoustic characteristics of 32 male great tits, each of which had a repertoire of between three and nine distinct song types. Mean song frequencies varied considerably between individual birds. The average minimum frequency ranged from 2.82 to 3.77 kHz and was significantly correlated with ambient noise (multiple regression: $n = 32$, d.f. = 2, $F = 4.74$, $P = 0.017$), with regard to both amplitude level ($t = 3.02$, $P = 0.005$) and spectral distribution ($t = -2.0$, $P = 0.055$). Noisy territories were home to great-tit males whose songs had a high average minimum frequency. Birds in quiet territories sang more notes that reached the lowest frequencies

measured for the population (Fig. 1). It is possible that individuals with genetically predetermined song spectra could end up in matching territories with regard to noise spectra through a process of trial and error. But we know that great tits learn their song and that major adjustments occur in their breeding territory during interactions with neighbours⁹. Hence it is more likely that they learn to use a restricted range of their spectral capacity in response to frequency-dependent interference from local noise conditions — adjusting song to territory instead of territory to song.

Frequency use by great tits is also known to vary with sound-transmission properties in different habitats¹⁰, and correlations with natural ambient-noise spectra are found in other birds^{11,12}. Local song adjustment within the heterogeneous urban habitat, as we find here, might indicate that habitat-wide adjustment through song learning may also contribute to acoustic divergence at the population level^{12,13}.

Our findings show, to our knowledge for the first time, that human-altered environments might change the communication signals of a wild bird species⁵. The apparent song plasticity of great tits may represent a



general behavioural mechanism that allows more bird species to reproduce despite high noise levels. Species that lack such learning plasticity after dispersal to the breeding territory, or have no room for variation within the conspecific frequency range, might suffer from auditory masking. For those species, anthropogenic noise could affect breeding opportunities and contribute to a decline in species density and diversity.

Hans Slabbekoorn, Margriet Peet

Behavioural Biology, Institute of Biology Leiden, Leiden University, 2300 RA Leiden, The Netherlands
e-mail: slabbekoorn@rulsfb.leidenuniv.nl

1. Reijnen, R. et al. *J. Appl. Ecol.* **32**, 187–202 (1995).
2. Catterall, C. P. et al. *Biol. Conserv.* **84**, 65–81 (1998).
3. Clergeau, P. et al. *Condor* **100**, 413–425 (1998).
4. Forman, R. T. T. & Alexander, L. E. *Annu. Rev. Ecol. Syst.* **29**, 207–231 (1998).
5. Rabin, L. A. & Greene, C. M. *J. Comp. Psychol.* **116**, 137–141 (2002).
6. Skiba, R. J. *Ornithol.* **141**, 160–167 (2000).
7. Junker-Bornholdt, R. et al. *J. Ornithol.* **139**, 131–139 (1998).
8. Klump, G. M. in *Ecology and Evolution of Acoustic Communication in Birds* (eds Kroodsma, D. E. & Miller, E. H.) 321–338 (Cornell Univ. Press, New York, 1996).
9. McGregor, P. K. & Krebs, J. R. *Behaviour* **108**, 139–159 (1989).
10. Hunter, L. M. & Krebs, J. R. *J. Anim. Ecol.* **48**, 759–785 (1979).
11. Slabbekoorn, H. & Smith, T. B. *Evolution* **56**, 1849–1858 (2002).
12. Slabbekoorn, H. & Smith, T. B. *Phil. Trans. R. Soc. Lond. B* **357**, 493–503 (2002).
13. Hansen, P. *Anim. Behav.* **27**, 1270–1271 (1979).

Competing financial interests: declared none.

Fluid dynamics

Vortex rings in a constant electric field

Despite nearly two centuries of study, the response of colloidal suspensions to electric fields can be surprising, particularly if the suspended particles also interact with each other. We describe a previously unrecognized class of hierarchically organized dynamic patterns that arises in aqueous

charge-stabilized colloidal suspensions when electrohydrodynamic forces due to constant applied fields compete with gravity.

Our system (Fig. 1a) consists of an aqueous suspension of monodisperse colloidal silica spheres 3.0 μm in diameter (Bangs Labs, lot no. 4740), confined to a thin horizontal layer between a glass coverslip and a microscope slide. Both inner surfaces are coated with 10-nm-thick gold electrodes on 10-nm-thick titanium wetting layers. Although still optically thin, these electrodes have a resistivity

of less than 50 ohms per square and allow us to apply uniform vertical electric fields to the confined suspension. Once equilibrated in air to pH 5.5, the silica spheres develop a surface charge density of -0.4 mC m^{-2} (ref. 1). As their density is 2 g cm^{-3} , they sediment rapidly onto the lower electrode.

Ions in solution screen out the electric field for biases below 2.4 V. Sustained upward forces occur only at higher biases for which hydrolysis at the electrodes feeds steady-state ionic fluxes, which in turn exert electroviscous forces on the charged spheres^{2,3}. These fluxes are spatially uniform in the parallel plate geometry, so that the force on an isolated sphere is independent of height h . Consequently, well-separated spheres (less than 0.01 monolayer) rise straight to the upper electrode at biases high enough to overcome gravity.

Distortions in the ionic fluxes due to the charged spheres mediate long-range interactions between spheres², which can induce interesting cooperative behaviour. In suspensions forming more than a monolayer, the spheres are levitated by biases exceeding 2.6 V into hundreds of flower-like clusters such as the example in Fig. 1a, all floating at $h = 40 \mu\text{m}$. Each cluster consists of a toroidal vortex in which spheres travel downwards along the inside and return upwards along the outside, completing one cycle in a few seconds. Most are surrounded by diffuse circulating coronas extending outwards for tens of micrometres. The clusters' circulation is consistent with a central downward flux of hydronium ions surrounded by an upward-moving sheath of hydroxyl ions, the resulting charge separation being supported by the spheres' charges and excluded volumes^{2,3}. However, this does not seem to explain all of the clusters' features.

Many circulating clusters enclose colloidal crystals⁴, most of which are monolayers. These close-packed domains disperse immediately once the driving field is turned off. Consequently, each cluster seems to straddle a stagnation plane, quite unlike a conventional vortex ring⁵.

Individual clusters sometimes develop breathing-mode instabilities with periods of a few seconds. One oscillating cluster's radius of gyration is tracked in Fig. 1b, c. Remarkably, the cluster's crystalline core reforms with each cycle. The oscillations of neighbouring clusters do not become phase-locked, and steadily circulating clusters can coexist with oscillating clusters. Consequently, distortions in the ionic fluxes entrained by a cluster seem to be well localized.

Clusters drift freely across the field of view, and so do not form at defects in the electrodes. Substantially smaller silica spheres do not form rings, but instead rise directly to the upper electrode, where they form interfacial crystals^{2,6}, as do less dense polystyrene and poly-(methyl methacrylate) spheres. So we conclude that colloidal vortex rings emerge as a cooperative phenomenon

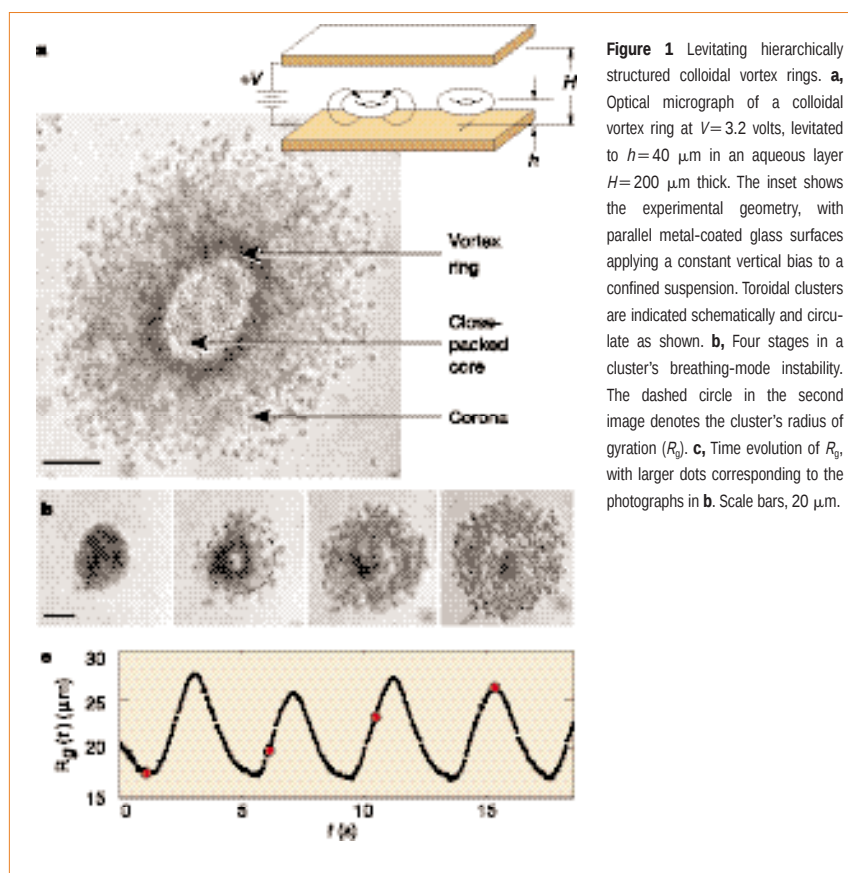


Figure 1 Levitating hierarchically structured colloidal vortex rings. **a**, Optical micrograph of a colloidal vortex ring at $V=3.2$ volts, levitated to $h=40 \mu\text{m}$ in an aqueous layer $H=200 \mu\text{m}$ thick. The inset shows the experimental geometry, with parallel metal-coated glass surfaces applying a constant vertical bias to a confined suspension. Toroidal clusters are indicated schematically and circulate as shown. **b**, Four stages in a cluster's breathing-mode instability. The dashed circle in the second image denotes the cluster's radius of gyration (R_g). **c**, Time evolution of R_g , with larger dots corresponding to the photographs in **b**. Scale bars, $20 \mu\text{m}$.

driven both by electrokinetic forces and by gravity. Understanding their origin could lead to new methods for creating opals of technologically relevant materials.

David G. Grier

The James Franck Institute, Institute for Biophysical Dynamics, and Department of Physics, University of Chicago, Chicago, Illinois 60637, USA
e-mail: d-grier@uchicago.edu

- Behrens, S. H. & Grier, D. G. *J. Chem. Phys.* **115**, 6716–6721 (2001).
- Trau, M., Saville, D. A. & Aksay, I. A. *Langmuir* **13**, 6375–6381 (1997).
- Squires, T. M. *J. Fluid Mech.* **443**, 403–412 (2001).
- Murray, C. A. & Grier, D. G. *Am. Sci.* **83**, 238–245 (1995).
- Shariff, K. & Leonard, A. *Annu. Rev. Fluid Mech.* **24**, U235–U279 (1992).
- Richetti, P., Prost, J. & Barois, P. *J. Physique Lett.* **45**, L1137–L1143 (1984).

Competing financial interests: declared none.

COMMUNICATIONS ARISING

Development rate

Modelling developmental time and temperature

The rate at which ectothermic animals develop is sensitive to temperature. Gillooly *et al.*¹ suggest that the magnitude of the effect of temperature on the rate of development is roughly the same at all life stages and for all species. In reality, different life stages and different species vary

considerably in terms of their sensitivity to changes in the thermal environment. This variability means that universal models based on cross-stage and cross-species comparisons, such as that proposed by Gillooly *et al.*, are of limited use in predicting how populations of animals will respond to temperature change.

The effect of temperature on development rate varies significantly during the course of development. For example, van't Hoff Q_{10} values for the duration of defined embryonic stages vary from 2.0 to 12.0 in the tailed frog *Ascaphus truei*² (interval Q_{10} ; Fig. 1a). (The van't Hoff Q_{10} is a simple indicator of thermal sensitivity based on the assumption of an exponential relationship between rate and temperature. A Q_{10} of 2.0 means that the rate doubles for each rise in temperature of 10°C ; a Q_{10} of 12.0 indicates that the rate increases 12-fold.)

Van't Hoff Q_{10} values for the duration of defined embryonic stages range from 1.9 to 8.2 in chinook salmon *Oncorhynchus tshawytscha*³ (Fig. 1b) and from 2.2 to 5.8 in the sea urchin *Strongylocentrotus droebachiensis*⁴ (Fig. 1c). The degree of variation is similar if the slope function, α , of Gillooly *et al.*, rather than the van't Hoff Q_{10} , is used to estimate thermal sensitivity. For example, values of α for the tailed frog range from -0.075 to -0.27 , equivalent to Arrhenius Q_{10} values of 2.1 to 14.9 (Arrhenius

Allometry

How reliable is the biological time clock?

A theoretical framework that describes the effect of temperature and body size on biological processes has long been sought. Gillooly *et al.*¹ use an elegant combination of the first principles of biochemical kinetics and allometric growth to derive a general definition of biological time by which, simply from a knowledge of how warm and how large an animal is, it is possible to calculate how long it will take to develop. Here I argue that their application of chemical kinetics in formulating the theory is problematical, undermining their claim that the theory fits the evidence.

Gillooly *et al.*¹ build on their equations for metabolism (B ; ref. 2) to predict that plots of $\ln(t/m^{1/4})$ against $T_c/(1 + (T_c/273))$, where t is the development time, T_c is the animal's temperature in degrees Celsius, and m is its mass, yield a universal straight line with a common slope $\alpha = -E/kT_0^2$, where E is the average activation energy for metabolic reactions, k is Boltzmann's factor and $T_0 = 273$ K. However, this relation is neither linear nor universal.

The temperature-dependence term in the equations used by Gillooly *et al.*^{1,2} is the van't Hoff-Arrhenius law: $B(T) = B(T_0)e^{E/k(T - T_0/T_0T)}$. Arrhenius showed that the Q_{10} is a simplification for conditions when T and T_0 are close together, a conclusion also reached by Gillooly *et al.*¹. When the van't Hoff-Arrhenius law was first applied to development time³, it became evident that the activation energy decreases as temperature increases. This is still apparent from Gillooly *et al.*¹: $\bar{E}$ decreases with temperature by a factor of no more than 49% (Fig. 1a).

Therefore, as k and T_0 are constants, the slope $\alpha = -\bar{E}/kT_0^2$ increases with temperature and the relationship is not linear. There is further temperature dependence beyond the 'exact' expression $e^{(-E/kT)(T_0/(1 + (T_0/T_0T)))}$ of Gillooly *et al.*¹. The overall pattern in some of their figures might seem to be linear owing to the combination of many curvilinear relationships (one for each species) in a single plot and to the previous adjustment of many of the original data to intervals of 5 °C (Fig. 1b).

The main objection to the application of chemical kinetics to biology⁴ is that vital action is arrested in the vicinity of 0 °C and not at -273 °C. Therefore, as the constants $B(T_0)$ and $a(T_0)$ should be zero for $T_0 = 273$ K, the equations of Gillooly *et al.*¹ have zero $a(T_0)$ in the denominator and cannot be solved. Ultimately, this is why the relationship is not linear: when body temperature approaches the biological zero, development ceases and the line becomes asymptotic to the y-axis. Gillooly *et al.*¹ do not solve the long-standing debate over whether reaction kinetics⁵ or diffusion of reactants^{4,6-8} is the limiting factor controlling temperature dependence in vital processes.

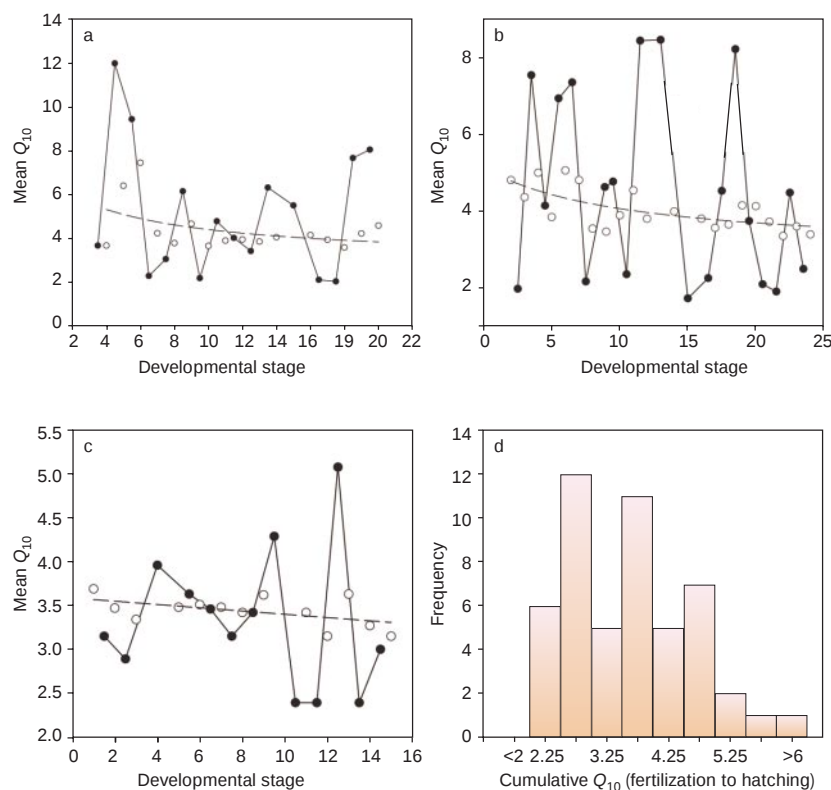


Figure 1 Stage- and species-specific variation in thermal sensitivity. Literature data are shown for **a**, tailed frog²; **b**, chinook salmon³; **c**, sea urchin⁴; and **d**, various teleost fish⁵. Development times (D) at various temperatures (T ; in °C) were fitted to the exponential model $D = ae^{bT}$, and mean van't Hoff Q_{10} values were calculated as e^{-10b} , where b is the slope of the regression equation. Interval Q_{10} values (filled circles) are for the duration of defined embryonic stages; cumulative Q_{10} values (open circles) are for the period from fertilization to a particular stage.

$Q_{10} = \exp(-10\alpha)$; Arrhenius $Q_{10} = \text{van't Hoff } Q_{10} \text{ at } 0^\circ\text{C}$.

This variation is unlikely to be the result of measurement errors. Stage durations are typically determined to an accuracy of within a few per cent of the reported time. Comparisons between species also indicate considerable variability. For example, van't Hoff Q_{10} values for the duration of the embryonic period (fertilization to hatching) of various species of teleost fishes ($n=50$) range between 2.1 and 6.9 (Fig. 1d)⁵; the equivalent range for Arrhenius Q_{10} values is about 2.5 to about 7.4.

Diversity is the salient characteristic in terms of temperature sensitivity at the stage and species level. It is necessary to look across stages and species to see any indication of uniformity. There are two principal reasons for this apparent narrowing of thermal responsiveness. One is a simple averaging effect. The net effect of stage-specific differences in thermal responsiveness becomes progressively less, the longer development proceeds (cumulative Q_{10} values in Fig. 1a–c). Mean response similarly converges in cross-species comparisons as more species are included in the analysis. The other reason is that evolutionary adaptation tends to reduce cross-species thermal sensitivity^{5,6}. For example,

rate functions of individual fish species typically show greater thermal sensitivity than would be predicted from cross-species comparisons, both during development⁵ and as adults⁶.

Changes in development rate have important implications for physiology (growth efficiency, for example) and ecology (such as predator–prey match/mismatch). Knowing the extent to which temperature affects development rate is therefore important, especially when viewed in the context of global warming. Universal models based on cross-stage and cross-species comparisons are unfortunately of little use in predicting how a particular species will respond. It will be necessary to examine each stage and species separately to come up with realistic predictions.

Peter Rombough

Department of Zoology, Brandon University,
Brandon, Manitoba R7A 6A9, Canada
e-mail: rombough@brandonu.ca

- Gillooly, J. F., Charnov, E. L., West, G. B., Savage, V. M. & Brown, J. H. *Nature* **417**, 70–73 (2002).
- Brown, H. A. *Comp. Biochem. Physiol. A* **50**, 397–405 (1975).
- Velsen, F. P. J. *Can. Data Rep. Fish. Aquat. Sci.* **626**, 19–20 (1987).
- Stephens, R. E. *Biol. Bull.* **142**, 132–144 (1972).
- Rombough, P. J. in *Global Warming: Implications for Freshwater and Marine Fish* (eds Wood, C. M. & McDonald, D. G.) 177–223 (Cambridge Univ. Press, 1997).
- Clarke, A. & Johnston, N. M. *J. Anim. Sci.* **68**, 893–905 (1999).

Combining temperature and body-mass dependence¹ is an interesting attempt to find out how the biological time clock works, but the clock seems to be imprecise. Data are presented on the development time for marine fish eggs in Fig. 2 of Gillooly *et al.*¹, who infer that the intercept is similar to that in their Fig. 1, but when these are plotted on the same y-axis scale they are strikingly different. Also, egg-development times for univoltine aquatic insects⁹ were excluded from their compilation. Taking an arbitrary temperature of 20 °C, the biological clock runs fast by 20 days for univoltine aquatic insects and slow by 5 days for marine fish (Fig. 1c). The intercepts are

therefore not similar, cautioning against the conclusion that the model of Gillooly *et al.* is common and invariant to all organisms.

Ángel López-Urrutia

Plymouth Marine Laboratory, The Hoe, Plymouth PL1 3DH, UK

e-mail: alop@pml.ac.uk

- Gillooly, J. F., Charnov, E. L., West, G. B., Savage, V. M. & Brown, J. H. *Nature* **417**, 70–73 (2002).
- Gillooly, J. F. *et al.* *Science* **293**, 2248–2251 (2001).
- Ludwig, D. *Physiol. Zool.* **1**, 358–389 (1928).
- Bělehrádek, J. *Protoplasma Monographien* **8** (1935).
- Crozier, W. J. *Proc. Natl. Acad. Sci. USA* **10**, 461–464 (1924).
- Bělehrádek, J. *Nature* **118**, 117–118 (1926).
- Heilbrunn, L. V. *Science* **62**, 268–269 (1925).
- Bělehrádek, J. *Nature* **118**, 478–480 (1926).
- Gillooly, J. F. & Dodson, S. I. *Freshwat. Biol.* **44**, 595–604 (2000).

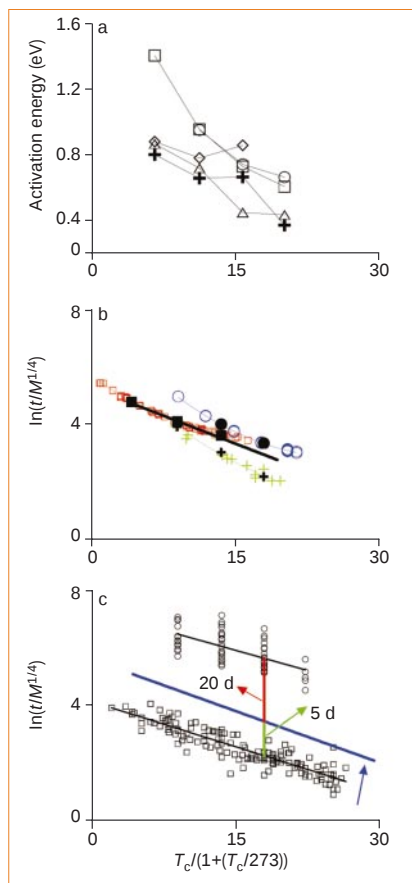


Figure 1 Effect of temperature on mass-corrected development time. **a**, Activation energy, E , calculated following Gillooly *et al.*¹, for different temperature intervals plotted against the interval's mean temperature. Lines link the average activation energy for species of amphibians (squares), fish (rhomboids), multivoltine aquatic insects (circles) and zooplankton (crosses, embryonic development; triangles, post-embryonic development). **b**, Colour symbols show the complete data sets, not adjusted to 5 °C intervals, for chinook salmon (squares), garfish (circles) and bream (crosses). Data sources from references in Gillooly *et al.*¹. Black symbols are the adjusted data for each of these species as used by Gillooly *et al.*¹; black line is their linear fit for fish (their Fig. 1b). **c**, 'Biological time clock' line (blue arrow) obtained by Gillooly *et al.*¹ for aquatic ectotherms and birds (their Fig. 3). Red and green lines show prediction error of the 'biological time clock' at 20 °C for non-diapause eggs of univoltine aquatic insects (circles) and marine fish (squares) incubated at different constant temperatures. Errors in days were calculated from egg masses of 0.01 and 0.9 mg, the average for each group from Gillooly *et al.*¹.

Gillooly *et al.* reply — Rombough asserts that our model¹ is of limited use in predicting the temperature dependence of developmental time, but he has used a model with the same two primary variables, size and temperature, for the same purpose². From his Fig. 1d above, he says that diversity among species is the salient feature for temperature dependence.

We replotted these data as the natural logarithm of mass-corrected incubation time against average incubation temperature in °C for fish with sizes and temperatures reported ($y = -0.18x + 4.34$; $r^2 = 0.72$; $n = 36$; Q_{10} ranges from 2 to 7; see supplementary information). The slope (-0.18) is close to the value predicted by our model (-0.14). Our model explains 72% of the variation, leaving only 28% to be explained by species differences, measurement errors and all other factors.

Rombough uses data on time intervals between stages to argue that the effect of temperature varies significantly during development. Most of this variation is measurement error because stages are defined arbitrarily, and times between them are short. Plotting cumulative times for three different stages from these data (Fig. 1) gives excellent fits and very similar slopes, showing that the temperature dependence remains nearly constant throughout development. Rombough actually presents additional support for our model.

López-Urrutia's main theoretical objection is that "vital action is arrested in the vicinity of 0 °C"; so our model is undefined. Biological activity ceases at around 0 °C because of a phase transition, the freezing of water. We consider this to be a separate process from molecular kinetics (which ceases at absolute zero). Therefore, by extrapolating the $a(T)$ curve for $T > 0$ to $T = 0$, we obtain a y-intercept, $a(0)$, that is always non-zero. López-Urrutia argues that activation energy (E) decreases systematically with temperature, and our figures only seem to be linear because we have averaged across temperatures for species. Plotting E for narrow temperature intervals, as in his Fig. 1a, is subject to measurement errors; the confidence intervals are infinite. Systematic changes in E with T should give curvilinear relationships in plots of $\ln(t/m^{1/4})$, as in his Fig. 1b, c or our Fig. 1. Although one species seems

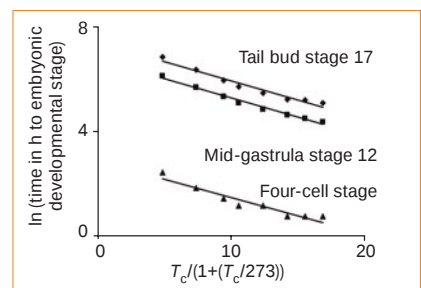


Figure 1 Plot of the natural logarithm of time to reach a specific embryonic development stage against incubation temperature for the tailed frog in Rombough's Fig. 1b (ref. 2). The slope of each line is close to the value predicted by our model (-0.13). For the tail-bud stage 17, $y = -0.15x + 7.4$ and $r^2 = 0.96$; for the mid-gastrula stage 12, $y = -0.15x + 6.7$ and $r^2 = 0.98$; and for the four-cell stage, $y = -0.14x + 2.8$ and $r^2 = 0.92$.

slightly curvilinear in his Fig. 1b, most species showed linear relationships across all our data.

Statistics provide a definitive answer for linearity. Curvilinear regression models (polynomials) did not give significantly better fits for any of our plots. Linearity is evident in López-Urrutia's Fig. 1c and in our Fig. 1, where linear fits account for 92–98% of the variation. Note that there is no averaging across temperatures for the fish in his Fig. 1c because each point represents a different species; all points in our Fig. 1 are for the same species.

In his Fig. 1c, López-Urrutia uses differences in intercepts, a , between fish in the wild, univoltine insects and the average for several taxonomic groups to question the reliability of the biological time clock. In our model, the coefficient a allows for variation in intercepts with metabolic rate, B_0 , and hence for differences in development times depending on which taxa, environmental conditions and developmental stage are measured. It is the $M^{-1/4}$ dependence on body size and $e^{-E/KT}$ temperature dependence in our equation (6), not the coefficient a , that defines the biological time clock. Note, however, that in López-Urrutia's Fig. 1c, the slopes and therefore the E values are nearly identical.

Both authors overlook our central message by focusing on temperature. We never claimed to have derived the Boltzmann factor or Q_{10} . Our model does provide a theoretical framework that combines the effects of size, temperature and stoichiometry to explain most of the variation in developmental rates across diverse environments and taxonomic groups.

James F. Gillooly*, Eric L. Charnov*, James H. Brown*, Van M. Savage†, Geoffrey B. West†

**Biology Department, University of New Mexico, Albuquerque, New Mexico 87131, USA*

e-mail: gillooly@unm.edu

†Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

- Gillooly, J. F., Charnov, E. L., West, G. B., Savage, V. M. & Brown, J. H. *Nature* **417**, 70–73 (2002).
- Rombough, P. J. in *Global Warming: Implications for Freshwater and Marine Fish* (eds Wood, C. M. & McDonald, D. G.) 177–223 (Cambridge Univ. Press, 1997).

Supplementary information accompanies this communication on Nature's website.

El Niño/Southern Oscillation and tropical Pacific climate during the last millennium

Kim M. Cobb^{*†}, Christopher D. Charles^{*}, Hai Cheng[‡] & R. Lawrence Edwards[‡]

^{*} Scripps Institution of Oceanography, University of California-San Diego, La Jolla, California 92093, USA

[‡] Minnesota Isotope Laboratory, Department of Geology and Geophysics, University of Minnesota, Minneapolis, Minnesota 55455, USA

Any assessment of future climate change requires knowledge of the full range of natural variability in the El Niño/Southern Oscillation (ENSO) phenomenon. Here we splice together fossil-coral oxygen isotopic records from Palmyra Island in the tropical Pacific Ocean to provide 30–150-year windows of tropical Pacific climate variability within the last 1,100 years. The records indicate mean climate conditions in the central tropical Pacific ranging from relatively cool and dry during the tenth century to increasingly warmer and wetter climate in the twentieth century. But the corals also document a broad range of ENSO behaviour that correlates poorly with these estimates of mean climate. The most intense ENSO activity within the reconstruction occurred during the mid-seventeenth century. Taken together, the coral data imply that the majority of ENSO variability over the last millennium may have arisen from dynamics internal to the ENSO system itself.

Despite recent advances in the understanding of the life cycle of the ENSO phenomenon—advances that have enabled climate models to forecast the evolution of an El Niño event with as much as six months lead time—many aspects of its physics remain unknown. Current ENSO theory holds that the growth and subsequent decay of El Niño events every few years is best explained with a recharge oscillator¹. However, there is still no consensus on the origin of ENSO's observed irregularity², which has been attributed variously to the effects of random atmospheric noise^{3,4}, low-order chaos^{5–7}, and low-frequency climate fluctuations^{8–10}.

One of the largest uncertainties concerns the relationship between ENSO characteristics and changes in the background climate state, whether natural or anthropogenic. The occurrence of very strong El Niño events in 1982 and 1997 has intensified the debate surrounding ENSO variability. To some extent, successful long-term climate prediction rests on the issue of whether the unusual severity of these events was a consequence of rising global temperatures¹¹ or was simply representative of natural variability^{12,13}. Numerical climate models provide few constraints: when charged with predicting how ENSO will change under continued greenhouse forcing, they yield a broad range of possible results^{14–16}. Given that the mean state of the tropical Pacific climate system is known to vary over decadal to centennial timescales^{17,18}, and may be changing in response to continued greenhouse forcing^{9,19}, it becomes especially important to determine the natural range of tropical Pacific climate variability and the controls on this variability. Specifically, extended records of ENSO across a variety of background climate conditions are required to test ENSO theories and models.

Long-lived corals provide continuous, high-resolution reconstructions of tropical Pacific climate that supplement the instrumental record of climate from this key region. Modern coral records from the central tropical Pacific that extend several centuries rival the fidelity of the instrumental record during the twentieth century, and have yielded insights into the recent history of tropical Pacific climate variability on a variety of timescales^{20–23}.

More recently, researchers have turned to fossil corals (colonies that are not living at the time of collection) to probe the character-

istics of the tropical Pacific climate system in the more distant past^{24–26}. Two limitations have hampered the widespread development of this approach. First, compared to modern corals, which are ubiquitous throughout the tropics, fossil corals of any significant length are extremely rare, and thus it has been difficult to reproduce fossil-coral-based climate proxy records. Second, because fossil-coral records are typically several decades long, they probably underestimate the full range of tropical Pacific climate variability in a given period. Even so, fossil corals represent the most direct source of monthly resolved records of tropical Pacific climate before about AD 1600. This unique capacity prompts us to pursue methods that might overcome the limits of the archive.

Here we generate multi-century, monthly resolved records of tropical Pacific climate variability over the last millennium by splicing together overlapping fossil-coral records from the central tropical Pacific. These precisely dated, well-reproduced records allow us to characterize the range of natural variability in the tropical Pacific climate system with unprecedented fidelity and

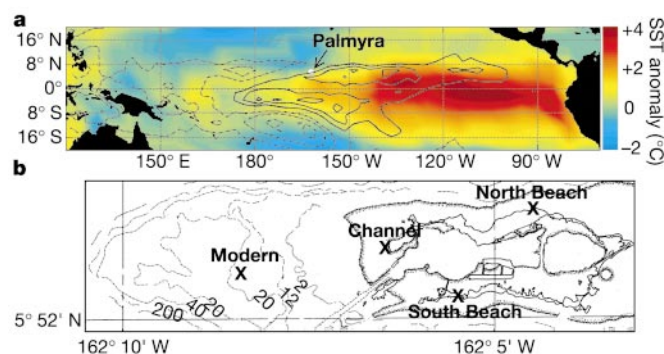


Figure 1 Maps of the study site. **a**, Map of the tropical Pacific showing location of Palmyra Island with respect to SST⁴⁵ (°C) and rainfall anomalies⁴⁶ (contour interval = 10 cm month⁻¹; solid (dashed) lines enclose positive (negative) anomalies) during the peak of the 1982–83 El Niño event. **b**, Map of Palmyra Island showing coral collection sites. Depth is contoured in metres. The modern coral was collected at the 'Modern' site, whereas the fossil corals were collected at the other three sites.

[†] Present address: California Institute of Technology, MC 100-23, Pasadena, California 91125, USA.

detail. Collectively, the records document a wide range of ENSO characteristics, along with substantial decadal-scale variability and more subtle centennial-scale fluctuations in tropical Pacific climate over the last millennium.

Splicing together fossil-coral records

We apply oxygen isotopic analysis to a large collection of fossil corals from Palmyra Island (6° N, 162° W) to construct proxy records of tropical Pacific climate variability during the last millennium (see Supplementary Methods). Climate at Palmyra is dominated by ENSO variability, manifest as warmer, wetter conditions that persist during El Niño events and cooler, drier conditions during La Niña events (Fig. 1a). Positive sea surface temperature (SST) and rainfall anomalies that occur at Palmyra during an El Niño event both result in lowered (more negative) coral $\delta^{18}\text{O}$, while the converse is true during a La Niña event. Thus, Palmyra corals are sensitive recorders of regional-scale ENSO activity. For example, a $\delta^{18}\text{O}$ record from a modern coral (*Porites lutea*) at Palmyra shares 72% of its interannual variance with the Niño3.4 index (Fig. 2), demonstrating that coral records from this site provide reliable proxies of ENSO activity, at least over the twentieth century²³. Our ability to reconstruct ENSO from fossil corals at this site rests in part on the assumption that the spatial patterns of ENSO have not changed significantly over the last millennium.

During two field excursions to Palmyra Island, we recovered dozens of 30–90-yr-long fossil-coral sequences from large, intact, *Porites* coral heads that were scattered on three different ocean-facing beaches (Fig. 1b). The Palmyra fossil-coral fields resemble storm-derived deposits that have been documented on other islands in the tropical Pacific²⁷, including nearby Christmas Island. The absence of extensive bioerosion in the Palmyra fossil corals supports the likelihood that wave activity from large, infrequent tropical storms transported the corals onto the beaches either before or shortly after their death, as the skeletons of dead corals are subject to rapid bio-erosion if they remain under water.

The numerous overlapping fossil coral sequences allow for the possibility of concatenating individual sequences that are detailed enough to resolve interannual variability into longer, well-reproduced sequences that span several centuries of tropical

Pacific climate. High-precision, accurate U/Th dates (generally ± 5 – 10 yr)²⁸ serve as guides for overlapping corals, but high coral-to-coral reproducibility of the $\delta^{18}\text{O}$ -based climate proxy records is required to make a definitive match between coral $\delta^{18}\text{O}$ sequences.

We use a young fossil coral that overlaps the modern coral during the early twentieth century to test both the U/Th-dating accuracy and the fidelity of young fossil coral $\delta^{18}\text{O}$ records. Five U/Th dates from the young fossil coral suggest that the beginning of the fossil-coral sequence falls between AD 1900 and 1915 (ref. 28). A firm match between the modern and fossil coral $\delta^{18}\text{O}$ records occurs when the fossil coral starts at AD 1915 (Fig. 3). The $\delta^{18}\text{O}$ reproducibility between the modern and fossil coral is highest in the interannual band (the corals share 83% of their 2–7-yr variance). In other words, the fossil coral and the modern coral contain similar records of ENSO variability throughout the interval of overlap.

We apply the same approach to overlap three coral $\delta^{18}\text{O}$ records from the seventeenth century and five coral $\delta^{18}\text{O}$ records from the fourteenth and fifteenth centuries (Fig. 4a and b, respectively). Two to five U/Th dates from each coral support the multi-coral $\delta^{18}\text{O}$ composites shown in Fig. 4 (see Supplementary Fig. S1). As in the twentieth-century case, each set of overlapping corals contains similar patterns of interannual $\delta^{18}\text{O}$ variability (correlation coefficients for overlapping 2–7-yr bandpassed records range from 0.67 to 0.87). Together, the overlapping corals provide strong evidence that ENSO variability is the dominant source of $\delta^{18}\text{O}$ variability in the Palmyra fossil-coral $\delta^{18}\text{O}$ records.

With the exception of the fourteenth–fifteenth century corals, whole-coral $\delta^{18}\text{O}$ offsets for overlapping sets of corals are smaller than analytical uncertainty ($\pm 0.05\text{‰}$, 1σ). However, the mean $\delta^{18}\text{O}$ values for the five overlapping corals from the fourteenth–fifteenth centuries are spread across a 0.3‰ range (Fig. 4b). Inter-coral offsets of up to 0.4‰ have been observed in neighbouring modern corals elsewhere in the central Pacific²⁹, and may reflect slight differences in the corals' microenvironments (water depth, sun exposure, circulation regime, and so on). Nonetheless, the range of offsets observed in the five fourteenth–fifteenth century fossil corals implies that there is a $\pm 0.12\text{‰}$ (1σ) (or $\pm 0.5^\circ\text{C}$, if scaled to temperature alone) error associated with reconstructing mean climate conditions from a single fossil-coral $\delta^{18}\text{O}$ record.

The splicing approach reduces this source of error by increasing the number of independent realizations of the climate state during a given time interval (assuming that mean coral $\delta^{18}\text{O}$ values are normally distributed about a value that represents the true climate state). Consequently, the error of our mean climate estimates for the twentieth century (based on two corals), the seventeenth century (based on three corals), and the fourteenth–fifteenth centuries (based on five corals) are all significantly less than $\pm 0.12\text{‰}$, the error for single coral estimates. However, this $\pm 0.12\text{‰}$ error does apply to two single fossil-coral $\delta^{18}\text{O}$ records from the tenth and twelfth centuries that comprise the earliest portions of the Palmyra reconstruction.

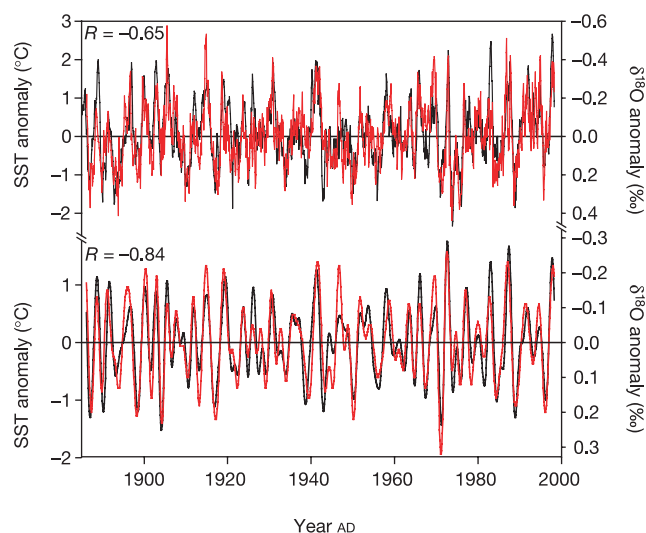


Figure 2 Comparison of coral-based and instrumental records of central tropical Pacific climate during the twentieth century. Top, Palmyra modern coral $\delta^{18}\text{O}$ anomalies (red; deseasoned, 30-yr highpass filtered) plotted with Niño3.4 SST anomalies (black; the average of SST anomalies from 5° N to 5° S, 120° W to 170° W)⁴⁷. Bottom, same as top but a 2–7-yr bandpass filter has been applied to the records shown above.

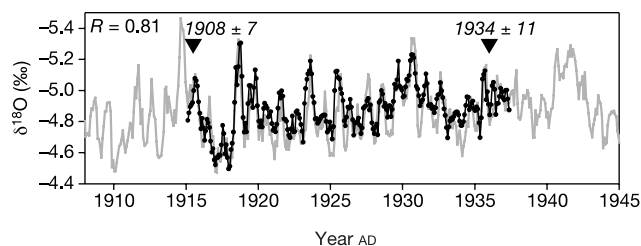


Figure 3 The $\delta^{18}\text{O}$ match between a young fossil coral from Palmyra (black) and the Palmyra modern coral (grey). Dating constraints for the fossil coral (in italics) come from multiple U/Th dates sampled from horizons marked by triangles²⁸.

The high degree of interannual reproducibility between the overlapping fossil-coral $\delta^{18}\text{O}$ records has two important implications for the reconstruction of tropical Pacific climate based on the Palmyra corals. First, it ensures that 'spliced' $\delta^{18}\text{O}$ records that are produced by joining the two longest coral $\delta^{18}\text{O}$ sequences at an arbitrary point can be viewed as continuous and accurate representations of climate variability. With current data, the splicing technique produces composite $\delta^{18}\text{O}$ records that are roughly twice as long as any given individual $\delta^{18}\text{O}$ record. Second, the inter-coral reproducibility lends credence to the interpretation of ENSO characteristics in individual fossil-coral $\delta^{18}\text{O}$ records, when overlapping corals are not available.

Tropical Pacific climate variability

We analyse the variability contained in monthly resolved coral $\delta^{18}\text{O}$ records from five intervals of the last millennium: AD 928–961, 1149–1220, 1317–1464, 1635–1703 and 1886–1998 (Fig. 5). The tenth- and twelfth-century sequences represent single fossil-coral $\delta^{18}\text{O}$ records, the fourteenth–fifteenth- and the seventeenth-century sequences are spliced fossil-coral $\delta^{18}\text{O}$ records, and the twentieth-century sequence is the modern coral $\delta^{18}\text{O}$ record.

Time-series analyses of the fossil-coral records reveal a range of ENSO frequencies and amplitudes exceeding that exhibited in the twentieth-century coral (Fig. 6a and b). For example, some seventeenth-century El Niño events rival the 1997 El Niño event in severity. ENSO activity in the seventeenth-century sequence is not only stronger (in terms of variance), but more frequent than ENSO activity in the late twentieth century. This conclusion holds for a variety of time-series analysis techniques, including spectral analysis and a range of different bandpass filters. On the other extreme, there

are 30-yr intervals during both the twelfth and fourteenth centuries when ENSO activity is greatly reduced relative to twentieth-century observations. Taken together, the fossil corals portray a highly variable ENSO over the last millennium whose amplitude and frequency changed markedly, in some cases over the course of a decade.

The fossil-coral data allow for a critical assessment of several theories that have been proposed to explain ENSO variability. First, the data test the suggestion that changes in the mean state, whether through natural decadal-scale variability or greenhouse warming, may alter ENSO characteristics^{10,11}. The relationship between ENSO variance and mean coral $\delta^{18}\text{O}$ is weak ($R = 0.43$), mostly because ENSO variance changed significantly while mean coral $\delta^{18}\text{O}$ remained relatively stable during the fourteenth–fifteenth centuries (Fig. 6a and c, and Supplementary Fig. S2). Overall, the corals resolve a broad range of ENSO variances that cannot be explained by changes in the mean state.

An alternative explanation for the observed irregularity of ENSO is that noise in the climate system, most likely of atmospheric origin, interferes with the recharge oscillator that is thought to set ENSO's periodicity^{1,3,4}. If random noise is the dominant source of ENSO's variability, then ENSO indices should be stationary—that is, the statistics of the time series (its spectral properties and variance) should not change appreciably through time. However, two of the fossil-coral sequences—the records from the twelfth and the fourteenth–fifteenth centuries—exhibit significant changes in ENSO behaviour from decade to decade (Fig. 6a). A definitive test for non-stationary behaviour requires longer records than those available at present, but the coral records produced thus far suggest that large fluctuations in ENSO variance are fundamental to the physics

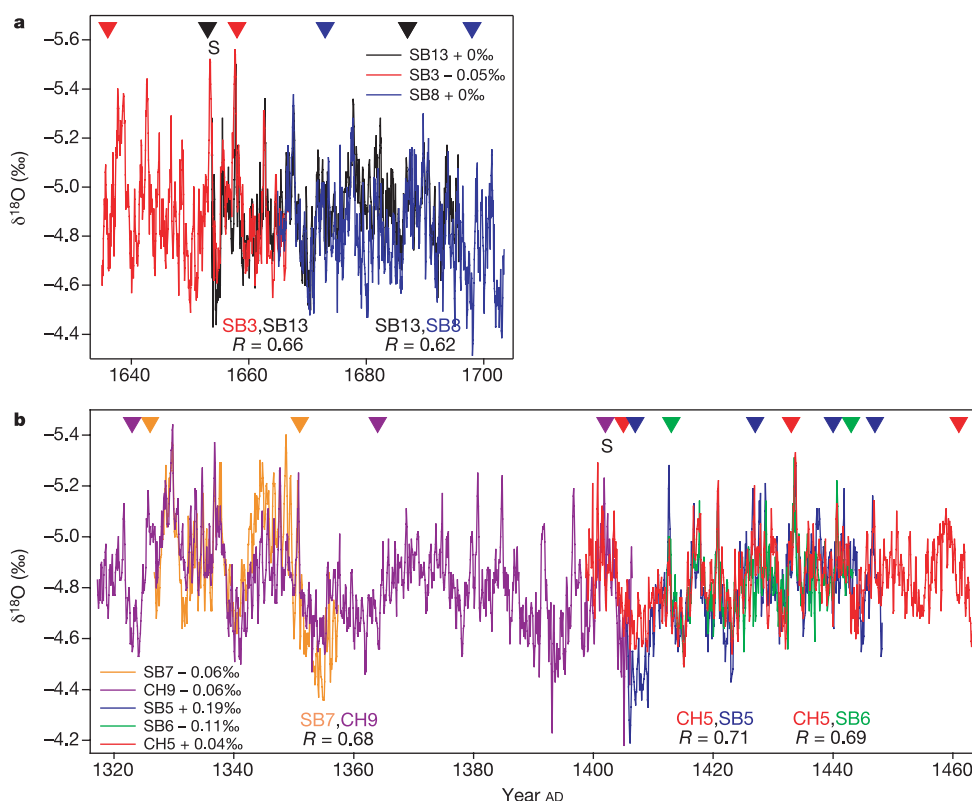


Figure 4 Overlapping fossil-coral $\delta^{18}\text{O}$ records. **a**, Three overlapping coral $\delta^{18}\text{O}$ records from the seventeenth century, shown with U/Th dating constraints (triangles; error is ± 5 – 10 yr for single U/Th dates, ± 5 yr for spliced record) and location of splice ('S'). The key reports the offsets in mean coral $\delta^{18}\text{O}$ that were applied prior to plotting. Correlation

coefficients for sets of overlapping coral $\delta^{18}\text{O}$ records are shown along the bottom. **b**, Same as **a** but for five overlapping coral $\delta^{18}\text{O}$ records from the fourteenth–fifteenth centuries. Axis scalings for the two plots are identical. SB and CH designations refer to samples collected at the South Beach and Channel sites, respectively.

of the phenomenon.

In fact, the behaviour exhibited by the coral records is reminiscent of ENSO 'regime changes' that occur in a variety of ENSO models whose variability is partially a product of chaos^{5–7,30}. The potential for regime-like changes in ENSO characteristics carries important implications for future climate changes under continued greenhouse forcing, because it allows for a nonlinear response of the global climate system to linear forcing.

Apart from ENSO variability, the Palmyra corals also resolve substantial decadal-scale variability that acts on a broad range of timescales from 8 to 30 yr. Power spectra of the records show no sign of a preferred periodicity for this low-frequency variability, which approaches a red noise continuum in the 150-yr-long sequence (not shown). However, it is clear from a comparison of Fig. 6a and c that decadal-scale variability is distinguishable from the decadal modulation of ENSO. This observation implies that the dynamics that underlie decadal-scale variability must be distinct from those of ENSO.

Given that coral $\delta^{18}\text{O}$ is a mixed SST and sea surface salinity signal, it is difficult to translate mean coral $\delta^{18}\text{O}$ into firm estimates of century-scale SST variability. However, if we assume that warmer SST is tightly coupled to anomalous convection in the central tropical Pacific on centennial timescales, as it is on interannual timescales, then the Palmyra corals provide unique constraints on the evolution of mean climate in the central tropical Pacific over the last millennium. Evidence for a tight SST–rainfall coupling on timescales longer than ENSO comes from the late-twentieth-century trend in coral $\delta^{18}\text{O}$. Recent estimates of central tropical Pacific warming since the 1970s are $\sim 0.8^\circ\text{C}^{19}$, which would correspond to a $\sim 0.14\text{‰}$ decrease in coral $\delta^{18}\text{O}$, using the SST/ $\delta^{18}\text{O}$ calibration from Fig. 2. The actual change in the coral $\delta^{18}\text{O}$ over this time period is about -0.30‰ , a roughly twofold amplification that must be ascribed, at least in part, to regional-scale freshening that accompanied central tropical Pacific warming. If twentieth-century SST–rainfall relationships apply throughout the last millennium,

then the Palmyra fossil corals should be sensitive recorders of mean climate change in the central tropical Pacific.

That said, mean coral $\delta^{18}\text{O}$ values for the twelfth, fourteenth–fifteenth, seventeenth and early twentieth centuries vary within a relatively narrow 0.14‰ range (or 0.6°C , if scaled to temperature alone). The only significant departures in mean coral $\delta^{18}\text{O}$ occur in the tenth century and late-twentieth-century sequences. It seems probable that the tenth century witnessed the coolest and/or driest conditions in the central tropical Pacific of the last 1,100 years, although this conclusion must be verified by additional tenth-century coral records. The late twentieth century, the period covered by dense instrumental climate data, represents the warmest, wettest interval of the last millennium.

Proxy–proxy comparisons

To build a complete interpretive framework for the Palmyra coral reconstruction, it would be necessary to compare the Palmyra fossil-coral ENSO results with proxy records of tropical Pacific origin that span the last millennium. Currently, multi-proxy reconstructions of tropical Pacific SST that extend back as far as the seventeenth century are the most likely candidates^{31–33}, but the earliest portions of these records are associated with large errors that render comparisons with the Palmyra corals questionable.

In lieu of high-resolution proxy records of tropical Pacific climate, we compare the Palmyra corals to reconstructions of Northern Hemisphere temperature that span the last millennium. The Mann *et al.*³⁴ multi-proxy reconstruction (hereafter MBH) is representative of most of the proxy-based Northern Hemisphere temperature reconstructions in that it broadly defines centuries of relatively warm and cool conditions, the 'Medieval Warm Period' (MWP) and the 'Little Ice Age' (LIA), respectively (Fig. 5). Crowley³⁵ reproduced several key features of the MBH reconstruction by combining volcanic and solar forcing in an energy-balance climate model. Solar irradiance changes have been linked to century-scale climate variability in a variety of tropical proxy records^{36,37}, while

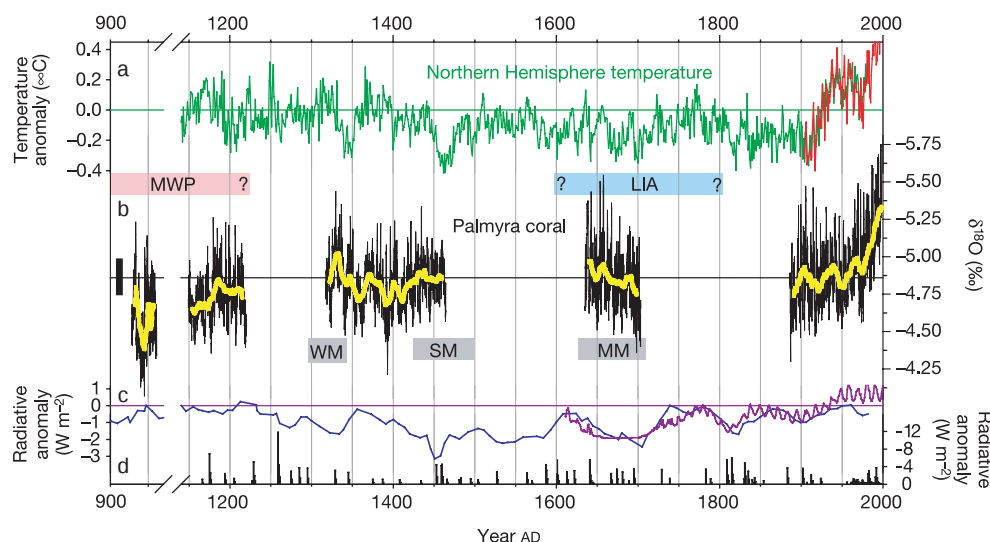


Figure 5 Comparison of proxy climate records and external forcing during the last millennium. **a**, The MBH Northern Hemisphere temperature reconstruction³⁴ (green) plotted with the Northern Hemisphere instrumental temperature record of ref. 48 (red). The green horizontal line denotes the mean of the MBH record for the period AD 1886–1975. **b**, The monthly resolved Palmyra coral $\delta^{18}\text{O}$ records (thin black line), shown with a 10-yr running average (thick yellow line). The black horizontal line represents the average of the Palmyra modern coral $\delta^{18}\text{O}$ for the period AD 1886–1975. The black vertical bar represents the $\pm 1\sigma$ error in mean coral $\delta^{18}\text{O}$ for single fossil corals (this error applies only to tenth- and twelfth-century sequences). The dating error is ± 10 yr for the tenth- and

twelfth-century sequences, ± 5 yr for the fourteenth–fifteenth- and seventeenth-century sequences, and $\pm <0.5$ yr for the twentieth-century sequence. **c**, Reconstruction of solar irradiance anomalies based on historical sunspot records (anomalies calculated with respect to the AD 1886–1975 mean)⁴⁹ (purple) plotted with ^{10}Be anomalies (a proxy for solar activity)⁵⁰ (blue), plotted as a 3-point running mean and scaled to the solar irradiance anomalies. **d**, Radiative forcing associated with volcanic eruptions recorded in ice cores (black)³⁵. The approximate timing and duration of the 'Little Ice Age' (LIA), the 'Medieval Warm Period' (MWP), and solar activity minima—the Maunder minimum (MM), the Spörer minimum (SP), and the Wolfe minimum (WM)—are marked by horizontal bars.

large volcanic eruptions have a measurable cooling effect on global average temperature³⁸. The links between the tropical Pacific climate system and solar (or volcanic) forcing over the last millennium are less clear. The combination of decreased irradiance and increased volcanism in the seventeenth century apparently did not result in strong immediate cooling of the central tropical Pacific. Likewise, increased irradiance and decreased volcanism did not result in immediate warming of the central tropical Pacific during the MWP. Furthermore, close inspection of the data (Fig. 5c and d versus Fig. 6) reveals little decadal-scale correspondence between the Palmyra reconstruction and the implied solar or volcanic forcing, either in terms of ENSO characteristics or variability in the mean state.

On the other hand, it can be argued that the east–west difference in SST across the tropical Pacific, not average SST, is a key determinant for global climate patterns. The Palmyra coral data, in combination with a handful of ENSO-sensitive proxy records, do suggest that the Pacific’s zonal SST gradient may have been larger during the MWP and smaller during the LIA. Higher mean $\delta^{18}\text{O}$ values in the tenth- and twelfth-century Palmyra corals hint at relatively cool and/or dry mean climate conditions in the central tropical Pacific, states that are both consistent with La Niña-like conditions. Cool conditions in the central tropical Pacific may have played a role in the severe, prolonged drought that lowered lake levels in Meso-America, the Sierra Nevada and Kenya^{39–41} early in the millennium, because central tropical Pacific cooling is typically associated with anomalously dry conditions in these areas during La Niña events⁴². Evidence for increased river runoff at about AD 1000 in northern South America⁴³, an area in which precipitation increases during La Niña events, also supports the inference of an increased Pacific zonal SST gradient during the MWP.

More direct evidence of tropical Pacific climate is available for the seventeenth century, when the Palmyra corals register the most intense ENSO activity of the reconstruction with little change in mean climate conditions. During this time, western Pacific corals

document cooler, drier conditions^{18,44} while river discharge off northern South America decreased⁴³. Both observations are consistent with a reduction in the tropical Pacific’s zonal SST gradient, and in that sense may be related to the frequent, intense El Niño events recorded in the Palmyra corals during the seventeenth century.

The most striking difference between the Palmyra reconstruction and the reconstructions of Northern Hemisphere temperature is the pattern of twentieth-century climate change, specifically the timing and structure of the warming trend. There are no *a priori* reasons to doubt the accuracy of either record—there are abundant instrumental data from the Northern Hemisphere extending to AD 1900, and while this is not the case for the central tropical Pacific, the salient features of the Palmyra modern coral $\delta^{18}\text{O}$ trend are reproduced in other modern coral $\delta^{18}\text{O}$ records^{21,22} from the central tropical Pacific. The mechanisms that are responsible for the stability of central tropical Pacific climate in the face of a $\sim 0.4^\circ\text{C}$ increase in Northern Hemisphere temperature during the early twentieth century may help explain why average central tropical Pacific climate remained relatively stable over the last millennium. On the other hand, the accelerated warming that took place in the central tropical Pacific during the late twentieth century implies very different forcing and response, in all probability related to the rise in greenhouse gases.

In summary, the Palmyra corals could provide critical tests of numerical climate models that are charged with predicting future climate change. Currently, such models are geared toward reproducing twentieth-century climate variability. Both short- and long-term predictions may probably improve if models could also reproduce the variability exhibited by the Palmyra fossil corals through the different boundary conditions of the last millennium. Perhaps most relevant in this respect is the fact that ENSO characteristics apparently changed quite markedly over the course of ten years, even in the absence of obvious external forcing. Therefore, it is appropriate to consider that similarly large, abrupt

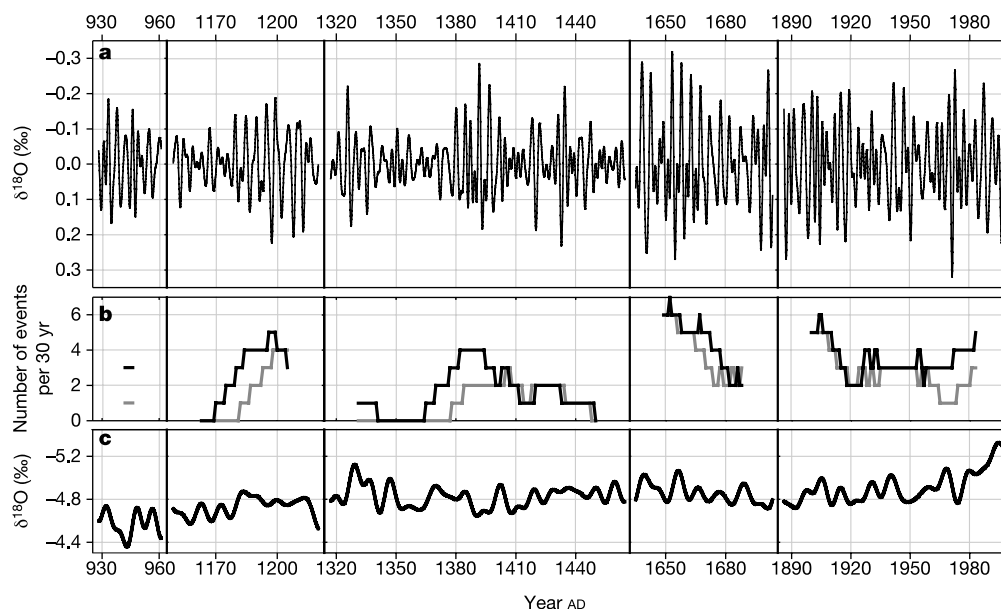


Figure 6 ENSO and lower-frequency components of the Palmyra coral $\delta^{18}\text{O}$ records. **a**, ENSO variability isolated by applying a 2–7-yr bandpass filter to the deseasoned monthly coral $\delta^{18}\text{O}$ anomaly data, plotted contiguously. **b**, An index of ENSO activity, defined as the number of El Niño (black) and La Niña (grey) events in a sliding 30-yr window. An El Niño (La Niña) event is defined by annual-mean $\delta^{18}\text{O}$ anomalies (computed

from the 2–7-yr bandpass filter series, centred on January) that are less than (greater than) -0.11‰ ($+0.11\text{‰}$). This threshold corresponds to one standard deviation of the modern coral’s 2–7-yr bandpassed record, and is roughly equivalent to Niño3.4 SST anomalies of 0.6°C , according to the calibration presented in Fig. 2. **c**, Lower-frequency climate variability isolated by applying an 8-yr lowpass filter to the coral $\delta^{18}\text{O}$ data.

changes in the character of tropical Pacific climate variability may occur in the future, with or without a trigger from continued greenhouse forcing. □

Received 17 March; accepted 30 May 2003; doi:10.1038/nature01779.

1. Jin, F. F. An equatorial ocean recharge paradigm for ENSO. 1. Conceptual model. *J. Atmos. Sci.* **54**, 811–829 (1997).
2. Neelin, J. D. *et al.* ENSO theory. *J. Geophys. Res.* **C 103**, 14261–14290 (1998).
3. Kleeman, R. & Power, S. B. Limits to predictability in a coupled ocean-atmosphere model due to atmospheric noise. *Tellus A* **46**, 529–540 (1994).
4. Graham, N. E. & White, W. B. The El Niño cycle: A natural oscillator of the Pacific ocean-atmosphere system. *Science* **240**, 1293–1302 (1988).
5. Jin, F. F., Neelin, J. D. & Ghil, M. El-Niño on the Devil's staircase—annual subharmonic steps to chaos. *Science* **264**, 70–72 (1994).
6. Tziperman, E., Stone, L., Cane, M. A. & Jarosh, H. El Niño chaos—overlapping of resonances between the seasonal cycle and the Pacific Ocean-Atmosphere Oscillator. *Science* **264**, 72–74 (1994).
7. Cane, M. A., Zebiak, S. E. & Xue, Y. in *Natural Climate Variability on Decade-to-Century Time Scales* (ed. Martinson, D. G. *et al.*) 442–457 (National Academy Press, Washington DC, 1995).
8. Gu, D. & Philander, S. G. H. Interdecadal climate fluctuations that depend on exchanges between the tropics and extratropics. *Science* **275**, 805–807 (1997).
9. Latif, M., Kleeman, R. & Eckert, C. Greenhouse warming, decadal variability, or El Niño? An attempt to understand the anomalous 1990's. *J. Clim.* **10**, 2221–2239 (1997).
10. Fedorov, A. V. & Philander, S. G. A stability analysis of tropical ocean-atmosphere interactions: Bridging measurements and theory for El Niño. *J. Clim.* **14**, 3086–3101 (2001).
11. Trenberth, K. E. & Hoar, T. J. The 1990–1995 El Niño–Southern Oscillation event: Longest on record. *Geophys. Res. Lett.* **23**, 57–60 (1996).
12. Rajagopalan, B., Lall, U. & Cane, M. A. Anomalous ENSO occurrences: An alternate view. *J. Clim.* **10**, 2351–2357 (1997).
13. Harrison, D. E. & Larkin, N. K. Darwin sea level pressure, 1876–1996: Evidence for climate change? *Geophys. Res. Lett.* **24**, 1779–1782 (1997).
14. Tett, S. Simulation of El Niño–Southern Oscillation-like variability in a global AOGCM and its response to CO₂ increase. *J. Clim.* **8**, 1473–1502 (1995).
15. Knutson, T. R., Manabe, S. & Gu, D. F. Simulated ENSO in a global coupled ocean-atmosphere model: Multidecadal amplitude modulation and CO₂ sensitivity. *J. Climate* **10**, 138–161 (1997).
16. Timmermann, A. *et al.* Increased El Niño frequency in a climate model forced by future greenhouse warming. *Nature* **398**, 694–697 (1999).
17. Zhang, Y., Wallace, J. M. & Battisti, S. ENSO-like interdecadal variability: 1900–93. *J. Clim.* **10**, 1004–1020 (1997).
18. Hendy, E. J. *et al.* Abrupt decrease in tropical Pacific Sea surface salinity at end of Little Ice Age. *Science* **295**, 1511–1514 (2002).
19. McPhaden, M. J. & Zhang, D. X. Slowdown of the meridional overturning circulation in the upper Pacific Ocean. *Nature* **415**, 603–608 (2002).
20. Cole, J. E., Fairbanks, R. G. & Shen, G. T. Recent variability in the Southern Oscillation—isotopic results from a Tarawa Atoll coral. *Science* **260**, 1790–1793 (1993).
21. Evans, M. N., Fairbanks, R. G. & Rubenstone, J. L. The thermal oceanographic signal of El Niño reconstructed from a Kiritimati Island coral. *J. Geophys. Res.* **C 104**, 13409–13421 (1999).
22. Urban, F. E., Cole, J. E. & Overpeck, J. T. Influence of mean climate change on climate variability from a 155-year tropical Pacific coral record. *Nature* **407**, 989–993 (2000).
23. Cobb, K. M., Charles, C. D. & Hunter, D. E. A central tropical Pacific coral demonstrates Pacific, Indian, and Atlantic decadal climate connections. *Geophys. Res. Lett.* **28**, 2209–2212 (2001).
24. McCulloch, M. *et al.* High-resolution windows into early Holocene climate: Sr/Ca coral records from the Huon Peninsula. *Earth Planet. Sci. Lett.* **138**, 169–178 (1996).
25. Tudhope, A. W. *et al.* Variability in the El Niño–Southern Oscillation through a glacial-interglacial cycle. *Science* **291**, 1511–1517 (2001).
26. Hughen, K. A., Schrag, D. P., Jacobsen, S. B. & Hantoro, W. El Niño during the last interglacial period recorded by a fossil coral from Indonesia. *Geophys. Res. Lett.* **26**, 3129–3132 (1999).
27. Scoffin, T. P. The geological effects of hurricanes on coral reefs and the interpretation of storm deposits. *Coral Reefs* **12**, 203–221 (1993).
28. Cobb, K. M., Charles, C. D., Cheng, H., Kastner, M. & Edwards, R. L. U/Th-dating living and young fossil corals from the central tropical Pacific. *Earth Planet. Sci. Lett.* **210**, 91–103 (2003).
29. Linsley, B. K., Messier, R. G. & Dunbar, R. B. Assessing between-colony oxygen isotope variability in the coral *Porites lobata* at Clipperton Atoll. *Coral Reefs* **18**, 13–27 (1999).
30. Timmermann, A. Changes of ENSO stability due to greenhouse warming. *Geophys. Res. Lett.* **28**, 2061–2064 (2001).
31. Evans, M. N., Kaplan, A., Cane, M. A. & Vintzileos, A. in *Inter-hemispheric Climate Linkages* (ed. Markgraf, V.) 53–72 (Cambridge Univ. Press, Cambridge, 2001).
32. Timmermann, A., Kaplan, A. & Cane, M. A. Pacific sea surface temperature field reconstruction from coral $\delta^{18}\text{O}$ data using reduced space objective analysis. *Paleoceanography* **17**, 10.1029/2000PA000590 (2002).
33. Mann, M. E., Bradley, R. S. & Malcolm, K. H. in *El Niño and the Southern Oscillation: Multiscale Variability and Global and Regional Impacts* (eds Diaz, H. F. & Markgraf, V.) 357–412 (Cambridge Univ. Press, Cambridge, 2000).
34. Mann, M. E., Bradley, R. S. & Hughes, M. K. Northern hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations. *Geophys. Res. Lett.* **26**, 759–762 (1999).
35. Crowley, T. J. Causes of climate change over the past 1000 years. *Science* **289**, 270–277 (2000).
36. Hodell, D. A., Brenner, M., Curtis, J. H. & Guilderson, T. Solar forcing of drought frequency in the Maya lowlands. *Science* **292**, 1367–1370 (2001).
37. Black, D. E. *et al.* Eight centuries of North Atlantic Ocean atmosphere variability. *Science* **286**, 1709–1713 (1999).
38. Robock, A. Volcanic eruptions and climate. *Rev. Geophys.* **38**, 191–219 (2000).
39. Hodell, D. A., Curtis, J. H. & Brenner, M. Possible role of climate in the collapse of Classic Maya civilization. *Nature* **375**, 391–394 (1995).
40. Stine, S. Extreme and persistent drought in California and Patagonia during Mediaeval time. *Nature* **369**, 546–549 (1994).
41. Verschuren, D., Laird, K. R. & Cumming, B. F. Rainfall and drought in equatorial east Africa during the past 1,100 years. *Nature* **403**, 410–414 (2000).
42. Dai, A. & Wigley, T. M. L. Global patterns of ENSO-induced precipitation. *Geophys. Res. Lett.* **27**, 1283–1286 (2000).
43. Haug, G. H., Hughen, K. A., Sigman, D. M., Peterson, L. C. & Rohl, U. Southward migration of the intertropical convergence zone through the Holocene. *Science* **293**, 1304–1308 (2001).
44. Corregge, T. *et al.* Little Ice Age sea surface temperature variability in the southwest tropical Pacific. *Geophys. Res. Lett.* **28**, 3477–3480 (2001).
45. Parker, D. E., Folland, C. K. & Jackson, M. Marine surface temperature: Observed variations and data requirements. *Clim. Change* **31**, 559–600 (1995).
46. Xie, P. P. & Arkin, P. A. Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Am. Meteorol. Soc.* **78**, 2539–2558 (1997).
47. Kaplan, A. *et al.* Analyses of global sea surface temperature 1856–1991. *J. Geophys. Res.* **C 103**, 18567–18589 (1998).
48. Jones, P. D. *et al.* Northern Hemisphere surface air temperature variations: 1851–1984. *J. Clim. Appl. Meteorol.* **25**, 161–179 (1986).
49. Lean, J., Beer, J. & Bradley, R. Reconstruction of solar irradiance since 1610—implications for climate change. *Geophys. Res. Lett.* **22**, 3195–3198 (1995).
50. Bard, E., Raisbeck, G., Yiou, F. & Jouzel, J. Solar irradiance during the last 1200 years based on cosmogenic nuclides. *Tellus B* **52**, 985–992 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank M. Moore and J. Arda for field assistance, T. Guilderson for preliminary radiocarbon dates, and A. Timmermann for comments on an early draft of the manuscript. We also thank the Khaled bin Sultan Living Ocean Foundation and The Nature Conservancy for financial and logistical support during two field excursions to Palmyra. K.M.C. was supported by a NSF graduate fellowship, and the work was supported by NOAA (C.D.C.) and NSF (R.L.E.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.M.C. (kcobb@gps.caltech.edu). The monthly Palmyra coral $\delta^{18}\text{O}$ data are available at (<http://www.ngdc.noaa.gov/paleo/pubs/cobb2003/>).

Genes that act downstream of DAF-16 to influence the lifespan of *Caenorhabditis elegans*

Coleen T. Murphy*, Steven A. McCarroll†, Cornelia I. Bargmann‡, Andrew Fraser‡, Ravi S. Kamath‡, Julie Ahringer‡, Hao Li* & Cynthia Kenyon*

* Department of Biochemistry and Biophysics, and † Department of Anatomy and Howard Hughes Medical Institute, University of California, San Francisco, California 94143-2200, USA

‡ Wellcome CRC Institute and Department of Genetics, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK

Ageing is a fundamental, unsolved mystery in biology. DAF-16, a FOXO-family transcription factor, influences the rate of ageing of *Caenorhabditis elegans* in response to insulin/insulin-like growth factor 1 (IGF-I) signalling. Using DNA microarray analysis, we have found that DAF-16 affects expression of a set of genes during early adulthood, the time at which this pathway is known to control ageing. Here we find that many of these genes influence the ageing process. The insulin/IGF-I pathway functions cell non-autonomously to regulate lifespan, and our findings suggest that it signals other cells, at least in part, by feedback regulation of an insulin/IGF-I homologue. Furthermore, our findings suggest that the insulin/IGF-I pathway ultimately exerts its effect on lifespan by upregulating a wide variety of genes, including cellular stress-response, antimicrobial and metabolic genes, and by downregulating specific life-shortening genes.

The recent discovery that the ageing process is regulated hormonally by an evolutionarily conserved insulin/IGF-I signalling pathway^{1–3} has provided a powerful entry point for understanding the causes of ageing at the molecular level. The nematode *C. elegans* lives for only a few weeks; however, animals carrying mutations that decrease insulin/IGF-I signalling, such as *daf-2* insulin/IGF-I receptor⁴ mutations, remain youthful and live more than twice as long as normal⁵. The insulin/IGF-I system also regulates reproduction^{5–7} and lipid metabolism⁴, as well as entry into a state of diapause called dauer⁸. The dauer is an arrested, long-lived juvenile form normally induced by food limitation and also by strong *daf-2* mutations⁸. The DAF-2 pathway regulates reproduction, lipid metabolism, dauer formation and ageing independently of one another^{9–12}. For example, whereas it acts during development to regulate dauer formation, it acts exclusively in the adult to influence ageing¹¹.

The DAF-2 pathway exerts its effects on the animal by influencing downstream gene expression, as the ability of *daf-2* mutations (*daf-2* (–)) to increase lifespan or produce other *Daf-2* (–) phenotypes depends on the activity of DAF-16 (refs 5–7, 11, 13), a FOXO-family transcription factor^{14,15}. In wild-type animals, the activity of DAF-16 is inhibited by a conserved phosphatidylinositol-3-OH kinase (PI(3)K)/protein kinase D (PDK)/Akt pathway in response to DAF-2 activity¹.

It should be possible to learn how insulin/IGF-I signalling influences ageing by identifying and characterizing the genes regulated by DAF-16. Some of these genes are predicted to encode or regulate downstream signals or hormones, because *daf-2* (and therefore presumably *daf-16*) functions cell non-autonomously^{12,16}: removing *daf-2* activity from subsets of cells can cause the entire animal to enter the dauer state, or to become a long-lived adult¹⁶. In addition, DAF-16 is predicted to influence expression of genes whose activities influence the ageing process directly. Animals with reduced DAF-2 pathway activity are resistant to heat and oxidative stress^{1,7,9,17,18}, which has suggested that an increased ability to prevent or repair oxidative damage increases lifespan. Consistent with this idea, overexpressing superoxide dismutase can extend the lifespan of *Drosophila*^{19,20} and yeast²¹. However, this hypothesis has

never been tested directly; for example, by asking whether stress response genes are required for the longevity of *daf-2* mutants. (An influential report was retracted recently²².)

In this study, we have identified genes that are regulated by DAF-16 and investigated their roles in the ageing process. To do this, we used microarray analysis to identify downstream genes, and then carried out a functional analysis of these genes using RNA interference (RNAi).

DNA microarray analysis

We identified genes whose expression changed in insulin/IGF-I pathway mutants using DNA microarrays containing approximately 93% of the predicted *C. elegans* open reading frames. We did this in two ways. First, we compared the transcriptional profiles of multiple alleles of long-lived *daf-2* and *age-1/PI(3)K* mutants to profiles of wild-type animals and *daf-16*; *daf-2* double mutants on the first day of adulthood. We grouped genes with similar expression patterns by hierarchical clustering of those with at least fourfold expression changes²³. This allowed us to identify genes that were upregulated or downregulated across the set of arrays. We also identified genes that were regulated in a highly consistent fashion, regardless of the degree to which their expression was changed²⁴ (Supplementary Table 1s).

In addition, we reduced *daf-2* and *daf-16* activity using RNAi, which phenocopies *daf-2* and *daf-16* mutants¹¹ (Fig. 1a). This allowed us to analyse the transcriptional profiles of isogenic and developmentally synchronous animals. We grew a sterile strain (*fer-15(b26)*; *fem-1(hc17)*) on bacteria expressing *daf-2* double-stranded (ds)RNA, both *daf-2* and *daf-16* dsRNA, or control bacteria, and collected the animals at intervals throughout adulthood (Fig. 1a). Because reducing the level of insulin/IGF-I signalling during early adulthood is sufficient to increase lifespan¹¹, we carried out an early adult time course (ten time points from 0–48 h of adulthood; Fig. 1a) to identify changes that occurred during this period. We also carried out a longer time course (ten time points from 0–192 h of adulthood) to identify changes that occurred as these animals began to age, but before a significant fraction died (Fig. 1a).

The early ageing transcriptome is largely unaffected

Because mutations in the insulin/IGF-I pathway slow the rate of ageing^{5,25,26}, we wondered whether reducing *daf-2* activity would slow the rate of all age-associated changes in gene expression. To investigate this, we compared the whole-transcriptome profiles of RNAi-treated animals at different ages (Fig. 1b, c). In the early adult time course, before the different strains began to differ morphologically²⁵ (0–48 h; Fig. 1b), we found that a subset of genes was expressed differently between animals exposed to *daf-2* RNAi and animals exposed to both *daf-16* and *daf-2* RNAi—we refer to these animals, which do not have long lifespans (Fig. 1a), as *daf-16(RNAi); daf-2(RNAi)* animals. These differences in expression persisted during the longer time course (0–192 h). During this period, the expression of many other genes changed as well; however, most of these age-dependent changes were not different between the *daf-2(RNAi)* and *daf-16(RNAi); daf-2(RNAi)* animals. This was surprising, as the tissue morphology of these animals differs significantly by the end of this period²⁵. Together these findings raised the possibility that the insulin/IGF-I pathway might influence ageing through a relatively small set of physiologically important targets that were differentially expressed even in young adults.

Two classes of downstream genes

We combined the data from our 60 microarrays into a single set and performed hierarchical clustering²³ (Fig. 2; see also Supplementary Information). We then focused on clusters that showed opposite expression profiles under *daf-2* (–) and *daf-16* (–) conditions. By examining a variety of mutants in multiple experiments and by performing two longitudinal studies, we were able to eliminate false positives caused by differences in developmental rates and by systematic errors. This approach revealed a relatively small number of differentially expressed *daf-2/daf-16*-dependent targets.

Two clusters were of particular interest. The first contained genes that were induced in DAF-2 pathway mutants and in *daf-2(RNAi)* animals but repressed in *daf-16(RNAi); daf-2(RNAi)* animals (class 1). These were candidates for genes that extend lifespan (Fig. 2; see also Supplementary Table 1s). The second cluster contained genes that displayed the opposite profile (class 2, Fig. 2; see also Supplementary Table 1s), and are candidates for genes that shorten lifespan. This approach identified genes previously thought to be regulated by DAF-16, such as the metallothionein homologue *mtl-1* (ref. 27) and the mitochondrial superoxide dismutase gene *sod-3* (ref. 28). We carried out polymerase chain reaction with reverse transcription (RT-PCR) of several RNA samples with *sod-3*- and *mtl-1*-specific primers, and found that expression of both was increased in *daf-2(RNAi)* animals (data not shown), confirming our microarray results.

A positive feedback loop amplifies DAF-2 pathway activity

In humans, reduced insulin receptor activity in the pancreas reduces insulin production. We found that gene expression of *ins-7*, which encodes an insulin/IGF-1-like peptide, was repressed in animals with reduced *daf-2* activity and elevated in animals with reduced *daf-16* activity. More than 35 insulin-like genes have been identified in the *C. elegans* genome^{29–31}, and 23 of these insulin-like genes were represented on our microarrays. A number of insulin-like peptides have been implicated in DAF-2 regulation^{30–32}. To investigate whether *ins-7* might function as a DAF-2 agonist, we inhibited its activity using RNAi. We found that *ins-7* RNAi increased the lifespan of *daf-2* (+) animals significantly (Fig. 3a), but was unable to further extend the lifespan of long-lived *daf-2* (*mu150*) animals (see Supplementary Table 2b). Furthermore, *ins-7* RNAi increased the frequency of dauer formation (Fig. 3b). Thus INS-7 behaved as expected for a DAF-2 agonist.

The regulatory properties of *ins-7* suggest that it might contribute to the non-autonomy of *daf-2* function. In this model, if *daf-2* gene activity is removed from cells that normally express *ins-7*, the level of *ins-7* expression will fall, which in turn will lower the level of INS-7 available to activate the DAF-2 receptor present on wild-type cells (Fig. 3c). In addition, the regulatory properties of INS-7 might contribute to an interesting phenomenon that occurs in nature. When a population of wild-type juvenile animals is confronted with a diminishing food supply (or when temperature-sensitive *daf-2* mutants are grown at a semi-permissive temperature), some but not all individuals enter the dauer state. It is interesting that under these threshold conditions, one does not observe animals containing random mixtures of dauer and non-dauer cells. It is possible that the INS-7 positive feedback loop contributes to this cellular conformity. In this model, a downward or upward fluctuation in the level of INS-7 would be amplified, which in turn would bias all of the cells in the animals towards dauer or adult development, respectively.

Additional downstream signalling molecules

In addition to *ins-7*, a number of other genes that encoded potential signalling molecules were regulated by DAF-2 and DAF-16. One was a known *daf-2/daf-16*-regulated gene, *scl-1*, which encodes a putative secreted protein that promotes longevity³³. Furthermore, a large

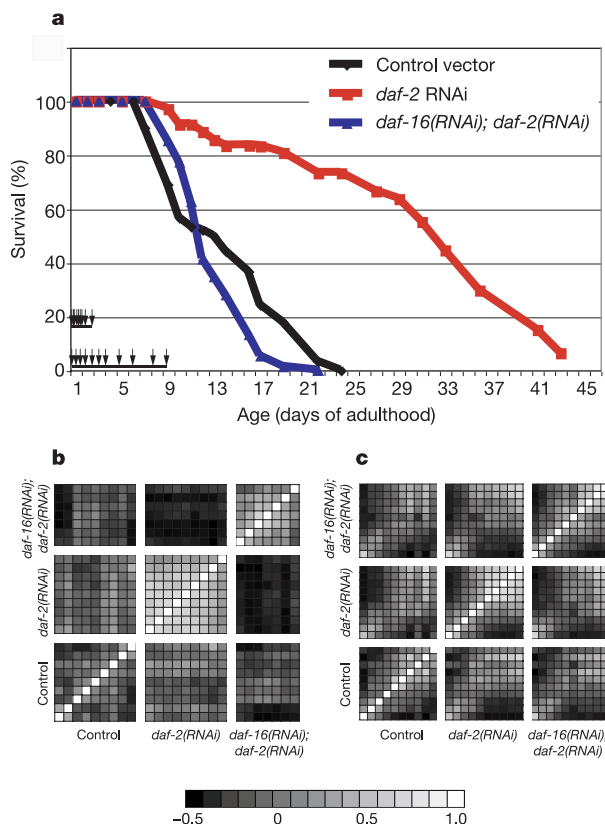


Figure 1 Effects of *daf-2* and *daf-16* RNAi on lifespan and the early ageing transcriptome. **a**, Lifespans of the RNAi-treated animals used for microarray analysis. Sterile *fer-15(b26); fem-1(hc17)* mutants were fed *daf-2*, *daf-2* and *daf-16*, or control RNAi bacteria. Arrows denote the times at which samples were taken for microarray analysis. **b**, **c**, Correlation coefficient analysis of arrays from the short time course (**b**; 0–48 h of adulthood) and long time course (**c**; 0–192 h of adulthood). Each time point's expression profile is expressed as a single value (see Methods), and the Pearson correlation coefficient describes its similarity to other time points (time points increase left to right and bottom to top). White indicates a perfect correlation, with increasing darkness indicating decreasing correlation.

number of class 1 (*daf-2* (–)-induced) genes encoded proteins that might potentially participate in the synthesis of a steroid or lipid-soluble hormone, including four cytochrome P450s, two estradiol-17- β -dehydrogenases, two alcohol/short-chain dehydrogenases, several esterases, two UDP-glucuronosyltransferases, and several *fat* genes known to function in fatty acid desaturation (Supplementary Table 1s). We investigated the functions of many of these genes and found that, in most cases, reducing their activities with RNAi shortened lifespan up to 20% (Fig. 4a, b). Together these findings suggest that the DAF-2 pathway may regulate multiple downstream signalling molecules. We also found that *gcy-6* and *gcy-18*, two receptor guanylate cyclases that are expressed in neurons³⁴, were repressed under *daf-2* (–) conditions (class 2). Inhibiting their activities lengthened the lifespan of *daf-2* (+) animals (Fig. 5a). Thus insulin/IGF-1 signaling may affect the animal's response to the environment.

A broad-based stress response increases longevity

A major goal of this study was to identify genes whose products directly influence ageing. We identified two prominent groups of functionally related genes that were candidates for direct effectors. The first group contained a wide variety of stress-response genes. In addition to *mtl-1* and *sod-3*, we found that expression of the catalase genes *ctl-1* and *ctl-2*, the glutathione-S-transferase gene *gst-4*, and the small heat-shock protein genes were all increased in animals with reduced *daf-2* activity and decreased in animals with reduced *daf-16* activity. We inhibited the activities of these genes with RNAi, and found that, in each case, the lifespans of *daf-2* mutants were shortened, generally between 10–20% (Fig. 4c, d and Table 1). Because DAF-16 also functions in the wild type to extend lifespan, inhibiting these genes would be predicted to shorten wild-type lifespan as well. This was often the case, although the magnitude of the effect was smaller than in *daf-2* (–) mutants (Supplementary Table 2s). Thus, each of these genes functions to promote longevity, probably by preventing or repairing oxidative and other forms of macromolecular damage.

An antimicrobial response lengthens lifespan

The second prominent set of potential lifespan effectors encoded antimicrobial proteins. *Caenorhabditis elegans* feeds on bacteria, and, at least under laboratory conditions, wild-type animals exhibit pharyngeal and intestinal bacterial packing as they age²⁵, and are ultimately killed by proliferating bacteria²⁵. *daf-2* mutants display reduced bacterial packing when compared with wild-type nematodes of the same age²⁵. We found that several genes encoding antibacterial lysosymes were induced in *daf-2* mutants, including two intestinally expressed genes, *lys-7* (C02A12.4) and *lys-8* (C17G10.5), which are also induced when *C. elegans* is infected with pathogenic *Serratia marcescens*³⁵. The saposin-like gene *spp-1* (T07C4.4), which has demonstrated antibacterial activity³⁶, was also upregulated in *daf-2* (–) animals. To test whether expression of these genes contributes to the longevity of *daf-2* mutants, we reduced the activities of several using RNAi. We found that these treatments shortened lifespan of *daf-2* mutants (Fig. 4e, f), indicating that these genes contribute to longevity.

Other *daf-2*/*daf-16*-regulated genes also influence lifespan

We found a number of other *daf-2*/*daf-16*-regulated genes with substantial effects on lifespan. For example, the vitellogenin (yolk protein/apolipoprotein-like) genes *vit-2* and *vit-5* were downregulated in *daf-2* (–) animals and upregulated in *daf-16* (–) animals, and we found that reducing their activities lengthened the lifespan of *daf-2* (+) animals (Fig. 5b). Several proteases and metabolic genes were also class 2 genes, including an aminopeptidase, a carboxypeptidase, an amino-oxidase, an aminoacylase, and *pep-2*, an oligopeptide transporter, as well as several F-box/cullin/Skp proteins (including *skr-8*, *skr-9* and *pes-2*) associated with ubiquitin-mediated protein degradation. Inhibition of several of these genes extended the lifespan of *daf-2* (+) animals (Table 2). This suggests that *daf-2* lifespan extension may involve turnover of specific proteins or metabolites. The glyoxylate cycle gene *gei-7* encoding isocitrate lyase/malate synthase, which is upregulated in dauers³⁷ and hibernating mammals³⁸, was upregulated in *daf-2* (–)

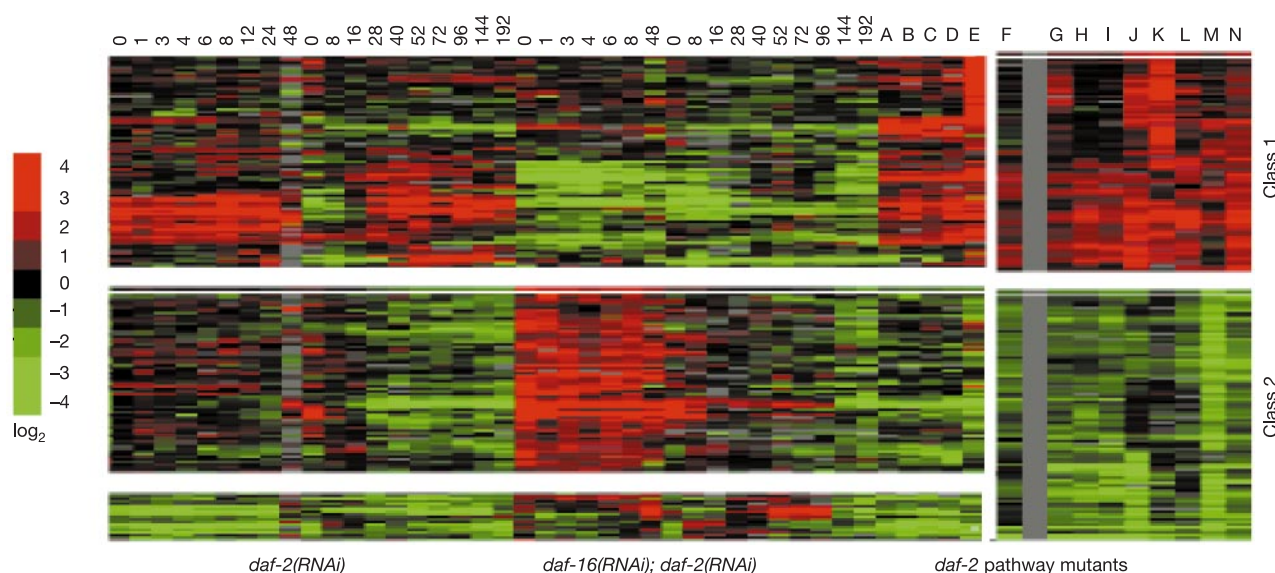


Figure 2 Class 1 genes are upregulated (red) with *daf-2* RNAi treatment and in *daf-2* pathway mutants, and downregulated (green) with *daf-16* RNAi treatment, whereas class 2 genes are upregulated with *daf-16* RNAi treatment and downregulated with *daf-2* RNAi treatment and in *daf-2* pathway mutants. Time points from the short and long time courses of *daf-2* and *daf-16* RNAi treatments and day 1 adult mutant comparisons are shown in hours along the top. (See Supplementary Materials for individual gene

expression profiles.) A, B, H, *fer-15* age-1; *fer-1* against *fer-15*; *fer-1*. C, D, G, I, *fer-15*; *daf-2(mu150)*; *fer-1* against *fer-15*; *fer-1*. E, *daf-16::gfp* in *daf-16(mu86)*; *daf-2(e1370)* against *daf-16(mu86)*; *daf-2(e1370)*. F, *daf-2(e1368)*; *fer-15*; *fer-1* against *fer-15*; *fer-1*. J, *daf-16::gfp* in *daf-2(e1370)*; *daf-16(mu86)* against *daf-2(e1370)*; *daf-16(mu86)*. K, *daf-2(mu150)* against wild type. L, M, N, *daf-2(e1370)* against *daf-2(e1370)*; *daf-16(mu86)*.

animals. Inhibiting the function of this gene shortened the lifespan of *daf-2* (–) mutants substantially, while shortening wild type lifespan only slightly (D. Cristina and C.K., unpublished data). Thus this alternative metabolic pathway contributes to longevity. A large class of unknown genes containing a shared domain of unknown function (the DUF141 domain) was downregulated in

daf-2 mutants, and RNAi of these genes extended lifespan (Fig. 5c). One gene that is repressed in *daf-2* mutants and induced in *daf-16* mutants, C54G4.6, had a relatively large effect on lifespan (Fig. 5c). This gene shares homology with bacterial orfE/MAF inhibitor of septum formation proteins and with a human protein, ASMTL³⁹. Finally, several other class 2 genes that significantly extended lifespan shared no homology with known genes (Fig. 5d and Table 2; see also Supplementary Table 2s).

A new potential regulatory sequence

To identify potential transcription-factor binding sites, we searched in an unbiased way for common sequence patterns in the upstream regulatory regions of genes in each cluster using two different algorithms. We used the 'Mobydick' algorithm⁴⁰ to identify short sequences (words) whose statistical distribution suggests that they are meaningful informational units. In this analysis we used sequences taken from a 1-kilobase (kb) region upstream of each gene in the cluster; words that are over-represented in the cluster are candidate transcription-factor binding sites. We also used another algorithm that searches exhaustively for oligonucleotides over-represented in each cluster⁴¹. We found that the sequence T(G/A)TTTAC, which has been shown to be bound by DAF-16 *in vitro*⁴², was over-represented, suggesting that our set of genes includes many direct DAF-16 targets. Notably, this canonical site was present not only in the promoters of class 1 (*daf-2*-induced) genes, but also in the promoters of class 2 (*daf-2*-downregulated) genes (Tables 1 and 2). Thus, DAF-16 may both directly repress and activate gene expression. We also found that a new sequence, CTTATCA, scored highly in both algorithms. Both sequences were present in various combinations in the promoters of both the class 1 and class 2 genes (Tables 1 and 2). The existence of this new site suggests that DAF-16 may regulate its target genes in combination with an additional, as yet unidentified, factor.

Mechanisms that modulate the rate of ageing

Together these findings suggest that the regulation of ageing by the insulin/IGF-I pathway is achieved through a combination of global regulators, such as INS-7 and neuronal signalling components, and a wide variety of genes whose products may affect the ageing process directly. Several DAF-16 target genes that had significant effects on lifespan encoded new proteins, and it will be interesting to learn whether these genes act in unexpected ways to influence lifespan. In addition, many DAF-16 target genes encoded proteins predicted to protect cells from oxidative and other forms of stress. Thus our study provides strong support for the theory that genes that increase resistance to environmental stress contribute to longevity.

In addition, our findings revealed that the ability to ward off microbial infections contributes to the longevity of *C. elegans*, and that this ability is regulated by insulin/IGF-I signalling. Bacterial infections are a major cause of disease and death in elderly humans. Thus, it will be interesting to learn whether the human insulin or IGF-I systems regulate the susceptibility to bacterial infections by controlling the expression of antimicrobial genes.

It was particularly interesting to find that no single RNAi treatment, other than *daf-16* RNAi itself, completely suppressed the lifespan extension of *daf-2* mutants. This was true also when we used a mutant strain with increased RNAi sensitivity⁴³ (Tables 1 and 2; see also Supplementary Table 2s(m)). This result indicates that multiple effector genes, whose expression is coordinated by the DAF-2 pathway, probably act in a cumulative manner to influence ageing. Because by themselves most genes have a relatively small effect on lifespan, it would have been difficult to identify any particular one in a standard genetic screen. Thus this study demonstrates the power of functional microarray analysis for dissecting complex regulatory systems.

Longevity must have evolved not just once, but many times. Insect lifespans range from a few weeks to several years, and those of

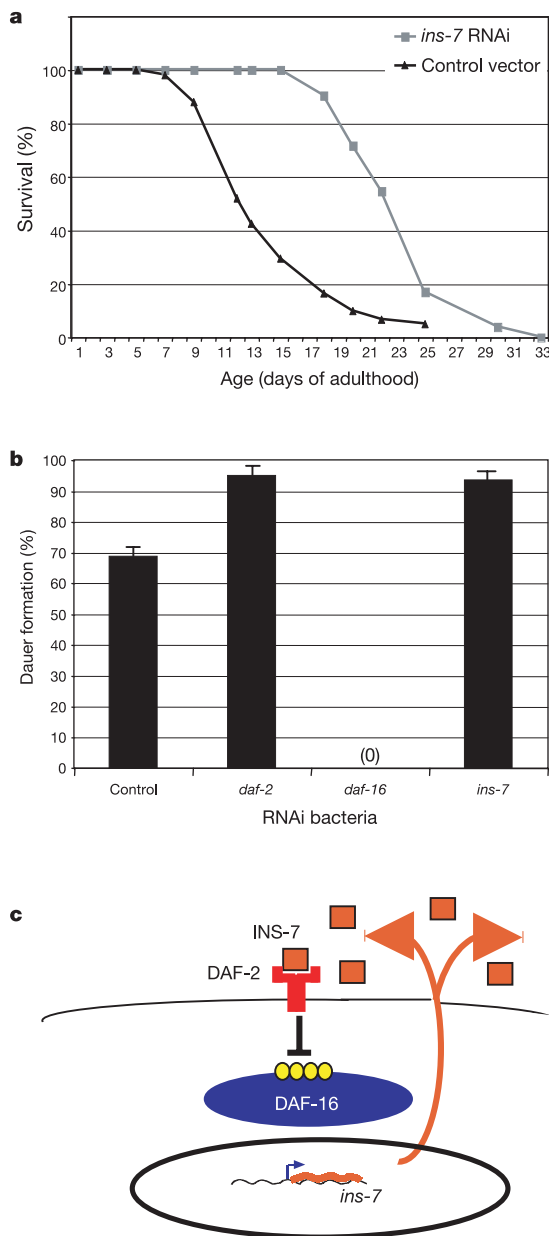


Figure 3 INS-7 behaves as a DAF-2 agonist, and is part of a positive feedback loop predicted to amplify DAF-2 pathway activity. **a**, *ins-7* RNAi extends the lifespan of the *daf-2*(+) RNAi-sensitive⁴³ strain *rrf-3(pk1426)*. *ins-7* RNAi also extends the lifespan of *fer-15*; *fem-1* animals (Supplementary Table 2s). **b**, *ins-7* RNAi increases the fraction of *daf-2*(e1370ts) mutants that become dauers. Parents were fed *ins-7* RNAi bacteria at 20 °C; their progeny were moved to 22.5 °C as eggs and observed 72 h later. **c**, Model of *ins-7* feedback regulation: when DAF-2 is active, DAF-16 activity is inhibited and *ins-7* is expressed, allowing further DAF-2 activation. When DAF-2 activity is reduced, DAF-16 is activated and *ins-7* expression is inhibited. (The expression of three other insulin-like genes changed in our microarrays: *ins-18* and *ins-2* were upregulated and *ins-21* was slightly downregulated in *daf-2*(–) animals.)

Table 1 **Class 1: Genes upregulated under *daf-2(-)* conditions**

Cosmid no.	Gene	Brief description	Per cent of vector control lifespan (experiment)				Canonical GTAAAT/cA	New CTTATCA
	<i>daf-16</i>		43.3(b)*	52.5(c)*	52.5(d)*	38.8(j)*	—	—
Y54G11A.5b	<i>ctl-2</i>	Peroxisomal catalase	54(a)*	92.9(b)†	89.6(c)¶	84.8(j)‡	1	3
T10B9.1	<i>dod-1</i>	Cytochrome P450 family, low similarity to mouse cytochrome P450 Cyp3a11	61.0(a)*	68.8(c)*			3	1
T27E4.8	<i>hsp-16.1</i>	Member of the <i>C. elegans hsp-16</i> family; identical <i>hsp-16.11</i>	71.3(d)†				0	0
C02A12.4	<i>lys-7</i>	Response to pathogenic bacteria; lysosyme/similar to N-acetylmuraminidase	72.6(a)*	92.9(c)¶	79.8(d)†	56.1(e)*	2	1
F28D1.3	<i>dod-2</i>	Thaumatococcus plant pathogenesis associated (PR) proteins, similar to F28D1.5	73.0(a)*	92.9(c)¶	92.2(j)¶		2	2
F38E11.2	<i>hsp-12.6</i>	Hsp20/alpha crystalline family, similar to alpha-B crystalline	75.8(a)‡	89.4(c)‡			3	2
K11G9.6	<i>mtl-1</i>	Metallothionein-related cadmium-binding protein	75.8(a)‡	89.4(c)¶			2	3
C05E4.9	<i>gei-7</i>	Malate synthase family/isocitrate lyase family	77.1(a)‡				5	3
C24B9.9	<i>dod-3</i>	Unknown protein	78.5(a)¶	99.7(c)¶			6	1
F32A5.5	<i>dod-4</i>	Aquaporin AQP; major intrinsic protein (MIP) family of transmembrane channels	78.7(d)*				1	4
T22G5.7	<i>dod-5</i>	Saposin type B	79.2(d)*	79.7(e)¶			4	1
F10D2.9	<i>fat-7</i>	Putative stearyl-CoA delta-9 fatty acid desaturase/polyunsaturated fatty acid biosynthesis	79.8(a)¶	88.2(c)¶	94.9(j)¶		3	1
T20G5.7	<i>dod-6</i>	Medtritrin-like ShK toxin	80.0(a)†	88.9(c)¶	82.3(d)†	72.3(e)¶	1	2
T27E4.9	<i>hsp-16.49</i>	Hsp20/alpha crystallin family, similar to alpha-B crystalline	81.0(d)†				4	1
C50B8.2	<i>bir-2</i>	Protein with two baculoviral inhibitor of apoptosis protein repeat (BIR) domains	81.1(c)*				5	1
T27E4.2	<i>hsp-16.11</i>	Member of the <i>C. elegans hsp-16</i> family; identical to <i>hsp-16.11</i>	81.4(d)†				1	0
T20G5.8		Medtritrin-like ShK toxin	84.3(d)‡	70.0(e)‡	97.5(j)¶		1	2
T07C4.4	<i>spp-1</i>	Saposin; similar to bactericidal amoebapores, may act as an antibacterial agent	84.3(d)§	77.0(e)¶			1	2
K11D2.2	<i>dod-7</i>	ASAH acid ceramidase; cholesteryl-glycine hydrolase, cleaves C-N non-peptide bonds in linear amides	85.4(c)‡				1	3
C06B3.4	<i>dod-8</i>	Estradiol 17b dh; short-chain dehydrogenase-reductase family oxidoreductases	87.6(c)§				4	1
Y54G11A.6	<i>ctl-1</i>	Cytosolic catalase	87.8(b)*	82.6(c)*	82.2(j)*		3	0
C46F4.2	<i>dod-9</i>	Acyl-CoA synthetase; high similarity to long-chain fatty acid-CoA ligase 4	87.8(c)‡				2	1
F43D9.4	<i>sip-1</i>	Hsp20/alpha crystallin family, moderately similar to <i>C. elegans</i> HSO-16 involved in heat shock response	88.4(c)¶				3	1
F11A5.12	<i>dod-10</i>	Short-chain dehydrogenase-reductase family, NAD- or NADP-dependent oxidoreductases	88.6(c)¶				1	3
C52E4.1	<i>gcp-1</i>	Cysteine protease expressed in the intestine	89.2(c)‡	92.9(j)¶			1	0
K12G11.3	<i>dod-11</i>	High similarity to <i>C. albicans</i> Adh1p, an alcohol dehydrogenase	89.6(a)¶	97.7(c)¶			3	2
R12A1.4	<i>ges-1</i>	Carboxylesterase expressed in gut cells	89.6(c)¶				1	3
C55B7.4	<i>dod-12</i>	Short branched chain acyl-CoA dehydrogenase (human ACADSB)	89.9(c)¶				0	1
H22K11.1	<i>asp-3</i>	Probable aspartyl protease and an orthologue of human cathepsin D	90.3(c)¶				3	0
Y46H3A.3	<i>hsp-16.2</i>	Strong similarity to <i>C. elegans</i> HSP-16 heat shock protein, Hsp20/alpha crystallin family	90.4(d)§				1	0
K07C6.4	<i>dod-13</i>	Cytochrome P450 family, low similarity to cytochrome P450 subfamily 2C polypeptide 8	90.6(b)*	84.9(c)‡			1	2
R03E9.1	<i>mdl-1</i>	MAD family of putative transcription factors, interacts with <i>C. elegans</i> MAX-1	91.1(d)¶				1	2
C08A9.1	<i>sod-3</i>	Manganese superoxide dismutase	92.6(b)¶	95.0(c)¶	83.2(j)*		6	2
K10B3.8	<i>gpd-2</i>	Glyceraldehyde-3-phosphate dehydrogenase	92.9(c)¶				4	0
K07E3.3	<i>dao-3</i>	Tetrahydrofolate dehydrogenase/cyclohydrolase catalytic domain, NAD(P)-binding domain	93.4(b)¶	83.9(c)‡	95.6(j)¶		5	2
T28B8.2	<i>ins-18</i>	Insulin-like protein of the type-beta subfamily; may be a ligand for the DAF-2 receptor	94.1(b)‡	88.1(c)¶			2	1
K12G11.4	<i>dod-14</i>	High similarity to <i>C. albicans</i> Adh1p alcohol dehydrogenase; Zn alcohol dehydrogenase family	95.0(b)¶	90.1(c)¶			4	1
AC3.7	<i>dod-15</i>	UDP-glucuronosyl, UDP-glucosyl transferase domains	95.1(c)¶				4	2
	<i>daf-2</i>		106.1(b)†	108.4(c)*	115.2(j)†		—	—
B0213.15	<i>dod-16</i>	Cytochrome P450, oxidation of arachidonic acid to eicosanoids; (mouse Cyp2j5)	119.0(a)‡	108.4(c)¶			1	2

The table is a summary of data from selected class 1 genes. Animals were treated with RNAi of selected genes and lifespans were compared to those of animals treated with control vector RNAi; experiments are briefly described below. "*dod*" stands for "downstream of DAF-16". (Detailed lifespan data are included in the Supplementary Information.) The number of canonical DAF-16 and new sequences in the 5 kb upstream of each gene is also shown. All experiments were performed with $n \geq 60$ animals. (a), *daf-2(mu150)*, 25 °C whole life; (b), *daf-2(mu150)* shifted from 20 °C to 25 °C at L3; (c), *daf-2(mu150)* shifted from 20 °C to 25 °C at L2; (d), *daf-2(mu150)* shifted from 20 °C to 25 °C at L2; (e), *daf-2(e1370)* shifted from 20 °C to 25.5 °C at L4; (j), *rrf-3(pk1426)*; *daf-2(e1370)* at 20 °C. * $P \leq 0.0001$; † $P \leq 0.001$; ‡ $P \leq 0.005$; § $P \leq 0.01$; ¶ $P \leq 0.05$; || $P > 0.05$.

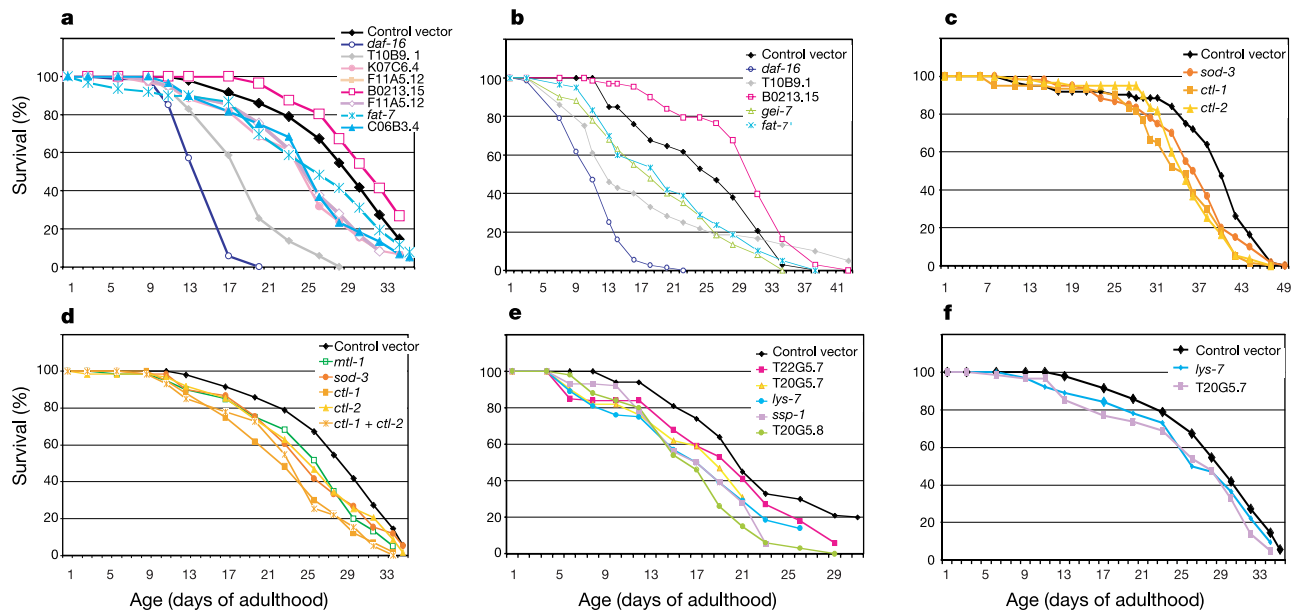


Figure 4 Lifespans of *daf-2* mutants fed dsRNA of class 1 genes. *daf-2(mu150)* mutants subjected to RNAi of metabolic and steroid and lipid synthesis genes (**a**, **b**) and oxidative stress genes (**c**, **d**). *daf-2(e1370)* (**e**) and *daf-2(mu150)* (**f**) mutants subjected to RNAi of antimicrobial genes. (Complete lifespan data are presented in Supplementary Table 2s.)

Table 2 Class 2: Genes downregulated under *daf-2(-)* conditions

Cosmid no.	Gene	Brief description	Per cent of vector control lifespan experiment			Canonical GTAAAt/cA	New CTTATCA
	<i>daf-2</i>		130.5(a)[†]	207 (b)*	191.5(c)*	—	—
K10D11.1	<i>dod-17</i>	DUF141 domain of unknown function, high similarity to uncharacterized <i>C. elegans</i> F55G11.8	133.7(c)*			4	3
C07B5.5	<i>nuc-1</i>	Endonuclease with strong similarity to <i>H. sapiens</i> DNase II: DNA degradation during apoptosis	130.4(c)*	101.9(d)		2	0
C54G4.6	<i>dod-18</i>	Maf-like protein family, inhibitors of septum formation, low similarity to uncharacterized <i>S. pombe</i> Spac3g6.03cp	129.2(a)*	132.4(b) [†]	126.4(c)*	0	2
ZK6.10	<i>dod-19</i>	Protein of unknown function	127.9(c)*			1	3
B0024.6	<i>gcy-6</i>	Putative guanylyl cyclase expressed in the ASEL neuron	126.1(c)*			2	2
B0554.6	<i>dod-20</i>	Protein of unknown function (DUF274) family, high similarity to uncharacterized <i>C. elegans</i> ZK6.11	123.0(c)*			3	5
C32H11.10	<i>dod-21</i>	DUF141 domain of unknown function, strong similarity to uncharacterized <i>C. elegans</i> C32H11.9	121.7(c) [†]			2	2
C04F6.1	<i>vit-5</i>	Vitellogenin; 170 kDa yolk protein	121.5(a) [†]	116.5(b)	109.7(c) [§]	1	1
T08G5.10	<i>mtl-2</i>	Protein of unknown function, has high similarity to uncharacterized <i>C. elegans</i> MTL-1	120.2(c) [†]	96.5(d)		1	1
F55G11.5	<i>dod-22</i>	DUF141 domain of unknown function, high similarity to uncharacterized <i>C. elegans</i> K10D11.2	118.1(c) [¶]			3	3
F49E12.2	<i>dod-23</i>	Protein of unknown function	116.5(c) [†]	101.1(d)		1	2
T22G5.2	<i>lbp-7</i>	High similarity to <i>C. elegans</i> LBP-5 (locomotory behaviour) ipocalin and cytosolic fatty-acid binding	114.4(c) [¶]	101.4(d)		0	4
K04E7.2	<i>pep-2</i>	Member of the proton-coupled oligopeptide transporter superfamily	113.7(c) [¶]			3	0
ZK1251.2	<i>ins-7</i>	Insulin-like protein of the type-beta subfamily	155.2(b)*	133.3(c)*		0	2
F56G4.2	<i>pes-2</i>	Unknown function, has very strong similarity to uncharacterized <i>C. elegans</i> F56G4.3	111.8(c)	124.5(d)		3	3
C08H9.5	<i>old-1</i>	Putative receptor tyrosine protein kinase; similar to human and <i>D. melanogaster</i> FGF receptor protein kinases	111.6(c) [¶]	109.6(d)		0	1
C32H11.12	<i>dod-24</i>	DUF141 domain of unknown function, high similarity to uncharacterized <i>C. elegans</i> C32H11.9	131.3(b)*	124.4(c)*		2	2
ZK896.8	<i>gcy-18</i>	Guanylate cyclase catalytic domain; receptor family ligand binding and protein kinase domain	125.5(b) [§]	124.7(c)*		3	1
C42D8.2	<i>vit-2</i>	Vitellogenin structural genes (yolk protein genes)	121.0(b) [¶]	124.4(c) [†]	1	1	2
	<i>daf-16</i>		79.3(c)*			-	-

The Table is the same as Table 1, except for class 2 genes (see Table 1 legend). (f), *fer-15(b26)*; *fer-1(hc17)* at 25 °C whole life; (g), *rrf-3(pk1426)* at 20 °C whole life; (h), *rrf-3(pk1426)* shifted to 25 °C at L2-L4, back to 20 °C; (i), *rrf-3(pk1426)* shifted to 25 °C at L2-L4, back to 20 °C.

**P* ≤ 0.0001; †, *P* ≤ 0.001; ‡, *P* ≤ 0.005; §, *P* ≤ 0.01; ¶, *P* ≤ 0.05; ||, *P* > 0.05.

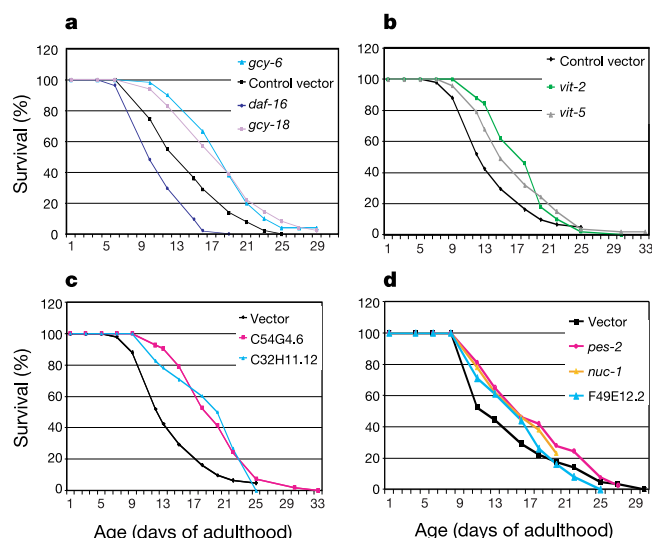


Figure 5 Lifespans of animals subjected to RNAi of indicated class 2 genes. *daf-2(+); rrf-3(pk1426)* mutants were grown on the different bacterial RNAi clones. (Complete lifespan data are presented in Supplementary Table 2s.)

mammals (and also birds) range from a few years to a century. Evolutionary theory postulates that lifespan is determined by the additive effects of many genes⁴⁴, consistent with our findings. The beauty of the insulin/IGF-I system is that it provides a way to regulate all of these genes coordinately. As a consequence, changes in regulatory genes encoding insulin/IGF-I pathway members or DAF-16 homologues could, in principle, allow changes in longevity to occur rapidly during evolution. Additional evolutionary flexibility may arise from the fact that the insulin and IGF-I system regulates longevity, reproduction, states of diapause and body size independently of one another^{2,7,11,12,45}. Thus, regulatory mutations that affect these traits differentially may allow evolving species to move into environmental niches that favour highly specific life history strategies.

Note added in proof: While this paper was in review, two independent studies of genes regulated by *daf-16* were published^{51,52}. We note that the gene called *ins-7* in ref. 52 may in fact be *ins-30*, which corresponds to the gene number cited in that report, ZC334.2. □

Methods

Microarray construction

We used Research Genetics 'GenePairs' primers for 18,455 predicted genes to amplify fragments by PCR from *C. elegans* N2 genomic DNA. PCR products were ethanol precipitated and size-confirmed before printing onto polylysine slides⁴⁶.

Strains

Mutations used in this study were: LG1, *daf-16* (*mu86*); LGII, *age-1* (*hx546*), *fer-15* (*b26*), *rrf-3* (*pk1426*)⁴³; LGIII, *daf-2* (*mu150*)²⁵, *daf-2* (*e1368*), *daf-2* (*e1370*); LGIV, *fem-1* (*hc17*); DAF-16::GFP strain, (*muEx 110* (*pKL199-2* (*daf-16::gfp* *daf-16b* (–)) + *pRF4(rol-6)*)); *daf-16* (*mu86*) I; *daf-2* (*e1370*) III¹⁷.

RNAi

Bacterial feeding RNAi experiments were carried out as described previously^{11,48}. We verified each clone from the RNAi library⁴⁸ by PCR and sequence analysis.

Caenorhabditis elegans growth and collection

A total of 30,000–50,000 nematodes were collected for each microarray sample. *daf-2* and *age-1* mutants were synchronized by hypochlorite treatment and L1 arrest, then grown to adulthood on 150 mm NG OP50 plates at 20 °C or 25 °C. Synchronized *fer-15* (*b26*); *fem-1* (*hc17*) animals were grown on RNAi bacteria at 25 °C and collected at the indicated time points (Fig. 1a); isopropyl-β-D-thiogalactoside was added on day 1 of adulthood and RNAi bacteria was supplemented as necessary. Nematodes were washed in M9, dissolved in Trizol (Gibco) and frozen in liquid nitrogen.

Microarray hybridizations

Standard techniques were used to obtain RNA (Trizol), messenger RNA (Oligotex, Qiagen), complementary DNA (reverse transcription) and Cy-dye-labelled cDNA⁴⁹; arrays were hybridized for 18 h at 63 °C, washed and scanned. One-half of each time course sample was added to a pool, and every Cy5-labelled sample was compared to this Cy3-labelled mixed reference. Mutant comparisons were done both directly and in a pooled comparison.

Significance analysis

After array normalization (see Supplementary Information), SAM analysis²⁴ was performed on data from nine mutant arrays (one-class response) to identify genes with small but consistent changes. In this set of arrays at a δ -value of 1.47, 70 upregulated and 100 downregulated genes were found to be significant (q -value = 0.0011194) with 0.6207 median false significant genes (Supplementary Table 1s).

Correlation coefficient analysis

We calculated a vector comprising the entirety of log ratio comparisons for all the genes with a valid signal at a single time point, to describe each array as a single value. We compared each array in the two time courses to all other arrays in that time course, and the Pearson correlation of the log base-two of these expression ratios was calculated. Five arrays from the set of 60 time points did not correlate with neighbouring time points, and were eliminated.

Cluster analysis

After normalization (see Supplementary Information), log transformation and quality confirmation through correlation coefficient analysis, data from 55 RNAi arrays and 5 mutant arrays were imported into Gene Cluster²³ for fold-cutoff analysis and hierarchical clustering. Genes were filtered to obtain only those that were present in 80% of the 60 arrays in the data set and which met a max–min of 4-fold, 8-fold or 16-fold criterion. A total of 7,380 genes met a 4-fold cutoff, 2,734 genes met an 8-fold cutoff and 1,280 genes met a 16-fold cutoff over the entire set of 55 RNAi arrays and five mutant arrays. The filtered set was hierarchically clustered, a self-organized map was constructed with 300,000 iterations, and the gene set was displayed in TreeView²³.

Upstream sequence analysis

The sequence 1 kb upstream of the translation start site of each open reading frame was assembled and subjected to two algorithms to search for potential transcription-factor binding sites. Exact repeats of length 14 or longer were removed before building the Mobydick⁴⁰ dictionary; words were screened by contrasting the frequency of occurrences in the cluster to that in the upstream regions of all the genes in the genome. We also searched for oligonucleotides over-represented in the cluster⁴¹. The occurrence of the identified sequences in the 5 kb upstream of each gene was then determined⁴⁰.

Survival analyses

Our lifespan analysis focused on a subset of genes whose expression profiles changed in opposite ways under *daf-2* (–) and *daf-16* (–) conditions. Genes were prioritized by fold expression change (Fig. 2) and by interesting gene function. The bacteria for 58 genes (Tables 1 and 2) were selected from the RNAi library⁴⁸. A total of 60–70 nematodes were used per experiment as described previously^{51,6}. The first day of adulthood was used as $t = 0$, and the log-rank (Mantel–Cox) method was used to test the null hypothesis (StatView 5.01, SAS Software). *fer-15* (*b26*); *fem-1* (*hc17*) animals were grown at 25 °C on RNAi bacteria and lifespans were measured at this temperature unless otherwise indicated. *daf-2* (*mu150*) nematodes were measured at 25 °C in one trial, and in subsequent tests were raised at 20 °C then shifted to 25 °C at L3. *rrf-3* (*pk1426*)⁴³ mutants were treated at 20 °C in one experiment; in subsequent experiments, the nematodes were shifted to 25 °C at L2 through young adulthood to induce sterility, and adult lifespan was measured at 20 °C. *rrf-3* (*pk1426*); *daf-2* (*e1370*) lifespan tests were done at 20 °C.

Dauer tests

daf-2 (*e1370*) nematodes were grown on RNAi bacteria at 20 °C, F₁ eggs were incubated at 22.5 °C, and animals were scored for dauer arrest 72 h later.

Received 17 December 2002; accepted 7 May 2003; doi:10.1038/nature01789.

Published online 29 June 2003.

1. Tatar, M., Bartke, A. & Antebi, A. The endocrine regulation of aging by insulin-like signals. *Science* **299**, 1346–1351 (2003).
2. Holzenberger, M. et al. IGF-1 receptor regulates lifespan and resistance to oxidative stress in mice. *Nature* **421**, 182–187 (2003).
3. Blüher, M., Kahn, B. B. & Kahn, C. R. Extended longevity in mice lacking the insulin receptor in adipose tissue. *Science* **299**, 572–574 (2003).
4. Kimura, K. D., Tissenbaum, H. A., Liu, Y. & Ruvkun, G. *daf-2*, an insulin receptor-like gene that regulates longevity and diapause in *Caenorhabditis elegans*. *Science* **277**, 942–946 (1997).
5. Kenyon, C., Chang, J., Gensch, E., Rudner, A. & Tabtiang, R. A *C. elegans* mutant that lives twice as long as wild type. *Nature* **366**, 461–464 (1993).
6. Larsen, P. L., Albert, P. S. & Riddle, D. L. Genes that regulate both development and longevity in *Caenorhabditis elegans*. *Genetics* **139**, 1567–1583 (1995).
7. Gems, D. et al. Two pleiotropic classes of *daf-2* mutation affect larval arrest, adult behavior, reproduction and longevity in *Caenorhabditis elegans*. *Genetics* **150**, 129–155 (1998).
8. Riddle, D. L. & Albert, P. S. In *C. elegans II* (eds Riddle, D. L., Blumenthal, T., Meyer, B. J. & Priess, J. R.) 739–768 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1997).
9. Guarente, L. & Kenyon, C. Genetic pathways that regulate ageing in model organisms. *Nature* **408**, 255–262 (2000).

10. Gems, D. & Partridge, L. Insulin/IGF signalling and ageing: seeing the bigger picture. *Curr. Opin. Genet. Dev.* **11**, 287–292 (2001).
11. Dillin, A., Crawford, D. K. & Kenyon, C. Timing requirements for insulin/IGF-1 signaling in *C. elegans*. *Science* **298**, 830–834 (2002).
12. Wolkow, C. A., Kimura, K. D., Lee, M. S. & Ruvkun, G. Regulation of *C. elegans* life-span by insulinlike signaling in the nervous system. *Science* **290**, 147–150 (2000).
13. Tissenbaum, H. A. & Ruvkun, G. An insulin-like signaling pathway affects both longevity and reproduction in *Caenorhabditis elegans*. *Genetics* **148**, 703–717 (1998).
14. Lin, K., Dorman, J. B., Rodan, A. & Kenyon, C. *daf-16*: An HNF-3/forkhead family member that can function to double the life-span of *Caenorhabditis elegans*. *Science* **278**, 1319–1322 (1997).
15. Ogg, S. *et al.* The Fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in *C. elegans*. *Nature* **389**, 994–999 (1997).
16. Apfeld, J. & Kenyon, C. Cell nonautonomy of *C. elegans daf-2* function in the regulation of diapause and life span. *Cell* **95**, 199–210 (1998).
17. Larsen, P. L. Aging and resistance to oxidative damage in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **90**, 8905–8909 (1993).
18. Lithgow, G. J., White, T. M., Hinerfeld, D. A. & Johnson, T. E. Thermotolerance of a long-lived mutant of *Caenorhabditis elegans*. *J. Gerontol.* **49**, B270–B276 (1994).
19. Orr, W. C. & Sohal, R. S. Extension of life-span by overexpression of superoxide dismutase and catalase in *Drosophila melanogaster*. *Science* **263**, 1128–1130 (1994).
20. Sun, J., Folk, D., Bradley, T. J. & Tower, J. Induced overexpression of mitochondrial Mn-superoxide dismutase extends the life span of adult *Drosophila melanogaster*. *Genetics* **161**, 661–672 (2002).
21. Longo, V. D., Liou, L. L., Valentine, J. S. & Gralla, E. B. Mitochondrial superoxide decreases yeast survival in stationary phase. *Arch. Biochem. Biophys.* **365**, 131–142 (1999).
22. Taub, J. *et al.* retraction: A cytosolic catalase is needed to extend adult lifespan in *C. elegans daf-C* and *clk-1* mutants. *Nature* **421**, 764 (2003).
23. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95**, 14863–14868 (1998).
24. Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA* **98**, 5116–5121 (2001).
25. Garigan, D. *et al.* Genetic analysis of tissue aging in *Caenorhabditis elegans*. A role for heat-shock factor and bacterial proliferation. *Genetics* **161**, 1101–1112 (2002).
26. Herndon, L. A. & Driscoll, M. Contributions of cell death to aging in *C. elegans*. *Results Probl. Cell Differ.* **29**, 113–129 (2000).
27. Barsyte, D., Lovejoy, D. A. & Lithgow, G. J. Longevity and heavy metal resistance in *daf-2* and *age-1* long-lived mutants of *Caenorhabditis elegans*. *FASEB J.* **15**, 627–634 (2001).
28. Honda, Y. & Honda, S. The *daf-2* gene network for longevity regulates oxidative stress resistance and Mn-superoxide dismutase gene expression in *Caenorhabditis elegans*. *FASEB J.* **13**, 1385–1393 (1999).
29. Gregoire, F. M., Chomiki, N., Kachinskas, D. & Warden, C. H. Cloning and developmental regulation of a novel member of the insulin-like gene family in *Caenorhabditis elegans*. *Biochem. Biophys. Res. Commun.* **249**, 385–390 (1998).
30. Pierce, S. B. *et al.* Regulation of DAF-2 receptor signaling by human insulin and *ins-1*, a member of the unusually large and diverse *C. elegans* insulin gene family. *Genes Dev.* **15**, 672–686 (2001).
31. Kawano, T. *et al.* Molecular cloning and characterization of a new insulin/IGF-like peptide of the nematode *Caenorhabditis elegans*. *Biochem. Biophys. Res. Commun.* **273**, 431–436 (2000).
32. Li, W., Kennedy, S. G. & Ruvkun, G. *daf-28* encodes a *C. elegans* insulin superfamily member that is regulated by environmental cues and acts in the DAF-2 signaling pathway. *Genes Dev.* **17**, 844–858 (2003).
33. Ookuma, S., Fukuda, M. & Nishida, E. Identification of a DAF-16 transcriptional target gene, *scl-1*, that regulates longevity and stress resistance in *Caenorhabditis elegans*. *Curr. Biol.* **13**, 427–431 (2003).
34. Yu, S., Avery, L., Baude, E. & Garbers, D. L. Guanylyl cyclase expression in specific sensory neurons: a new family of chemosensory receptors. *Proc. Natl Acad. Sci. USA* **94**, 3384–3387 (1997).
35. Mallo, G. *et al.* Inducible antibacterial defense system in *C. elegans*. *Curr. Biol.* **12**, 1209–1214 (2002).
36. Banyai, L. & Patthy, L. Amoebores homologs of *Caenorhabditis elegans*. *Biochim. Biophys. Acta* **1429**, 259–264 (1998).
37. Wang, J. & Kim, S. K. Global analysis of dauer gene expression in *Caenorhabditis elegans*. *Development* **130**, 1621–1634 (2003).
38. Davis, W. L., Goodman, D. B., Crawford, L. A., Cooper, O. J. & Matthews, J. L. Hibernation activates glyoxylate cycle and gluconeogenesis in black bear brown adipose tissue. *Biochim. Biophys. Acta* **1051**, 276–278 (1990).
39. Ried, K., Rao, E., Schiebel, K. & Rappold, G. A. Gene duplications as a recurrent theme in the evolution of the human pseudoautosomal region 1: isolation of the gene ASMTL. *Hum. Mol. Genet.* **7**, 1771–1778 (1998).
40. Bussemaker, H. J., Li, H. & Siggia, E. D. Building a dictionary for genomes: identification of presumptive regulatory sites by statistical analysis. *Proc. Natl Acad. Sci. USA* **97**, 10096–10100 (2000).
41. van Helden, J., Andre, B. & Collado-Vides, J. Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. *J. Mol. Biol.* **281**, 827–842 (1998).
42. Furuyama, T., Nakazawa, T., Nakano, I. & Mori, N. Identification of the differential distribution patterns of mRNAs and consensus binding sequences for mouse DAF-16 homologues. *Biochem. J.* **349**, 629–634 (2000).
43. Sijen, T. *et al.* On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell* **107**, 465–476 (2001).
44. Kirkwood, T. B. & Austad, S. N. Why do we age? *Nature* **408**, 233–238 (2000).
45. Clancy, D. J. *et al.* Extension of life-span by loss of CHICO, a *Drosophila* insulin receptor substrate protein. *Science* **292**, 104–106 (2001).
46. DeRisi, J. L., Iyer, V. R. & Brown, P. O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680–686 (1997).
47. Lin, K., Hsin, H., Libina, N. & Kenyon, C. Regulation of the *Caenorhabditis elegans* longevity protein DAF-16 by insulin/IGF-1 and germline signaling. *Nature Genet.* **28**, 139–145 (2001).
48. Fraser, A. G. *et al.* Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* **408**, 325–330 (2000).
49. DeRisi, J. *et al.* Genome microarray analysis of transcriptional activation in multidrug resistance yeast mutants. *FEBS Lett.* **470**, 156–160 (2000).
50. van Helden, J., Andre, B. & Collado-Vides, J. A web site for the computational analysis of yeast regulatory sequences. *Yeast* **16**, 177–187 (2000).
51. Lee, S. S., Kennedy, S., Tolonen, A. C. & Ruvkun, G. DAF-16 target genes that control *C. elegans* life-span and metabolism. *Science* **300**, 644–647 (2003).
52. McElwee, J., Bubb, K. & Thomas, J. H. Transcriptional outputs of the *Caenorhabditis elegans* forkhead protein DAF-16. *Aging Cell* **1**, 111–121 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank all the members of the Kenyon laboratory, as well as Z. Gitai, for critical review and discussion of this work, and A. Dillin, J. Lehrer-Graiwer, D. Cristina, B. Albindler, and V. Tenberg for assistance. We also thank J. DeRisi and the DeRisi laboratory for assistance and advice on the design and use of microarrays, as well as T. Kirkwood for discussions about evolution. S.A.M., from the laboratory of C.L.B., participated in the statistical analysis and the whole-transcriptome analysis; A.F., R.S.K. and J.A. contributed the RNAi clones; and H.L. participated in the promoter analysis. C.T.M. is a Bristol-Myers Squibb Fellow of the Life Sciences Research Foundation. This work was supported by grants from the NIA and the Ellison Foundation to C.K.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.K. (ckenyon@biochem.ucsf.edu). The data of the microarray experiments for *daf-2* mutants, *daf-2* RNAi (0–48 h) and *daf-2* RNAi (0–192 h) are deposited in ArrayExpress under accession numbers E-MEXP-7, E-MEXP-8 and E-MEXP-9, respectively.

Type II supernovae as a significant source of interstellar dust

Loretta Dunne*, Stephen Eales*, Rob Ivison†, Haley Morgan* & Mike Edmunds*

* Department of Physics and Astronomy, Cardiff University, 5 The Parade, Cardiff CF24 3YB, UK

† Astronomy Technology Centre, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK

Large amounts of dust ($>10^8 M_{\odot}$) have recently been discovered in high-redshift quasars^{1,2} and galaxies^{3–5} corresponding to a time when the Universe was less than one-tenth of its present age. The stellar winds produced by stars in the late stages of their evolution (on the asymptotic giant branch of the Hertzsprung–Russell diagram) are thought to be the main source of dust in galaxies, but they cannot produce that dust on a short enough timescale⁶ (<1 Gyr) to explain the results in the high-redshift galaxies. Supernova explosions of massive stars (type II) are also a potential source, with models predicting $0.2\text{--}4M_{\odot}$ of dust^{7–10}. As massive stars evolve rapidly, on timescales of a few Myr, these

supernovae could be responsible for the high-redshift dust. Observations^{11–13} of supernova remnants in the Milky Way, however, have hitherto revealed only $10^{-7}\text{--}10^{-3}M_{\odot}$ each, which is insufficient to explain the high-redshift data. Here we report the detection of $\sim 2\text{--}4M_{\odot}$ of cold dust in the youngest known Galactic supernova remnant, Cassiopeia A. This observation implies that supernovae are at least as important as stellar winds in producing dust in our Galaxy and would have been the dominant source of dust at high redshifts.

Over the past three decades, many searches for dust in supernova remnants have been made in the mid and far-infrared ($6\text{--}100\text{ }\mu\text{m}$). Remnants must be studied when they are young, before they have swept up large masses of interstellar material which makes it difficult to distinguish dust formed in the ejecta from that present in the interstellar medium (ISM) prior to the explosion. The handful of Galactic remnants which are both young and close enough (Cas A, Kepler and Tycho) have been studied with the Infrared Astronomical Satellite (IRAS) and the Infrared Space Observatory (ISO), but although dust at $100\text{--}200\text{ K}$ has been detected, the dust mass deduced is only $10^{-7}\text{--}10^{-3}M_{\odot}$, many orders of magnitude lower than the solar mass quantities predicted^{11–13}. The formation of dust in the recent supernova 1987A has been implied indirectly from the fading of the silicate line and increase in $10\text{ }\mu\text{m}$ emission¹⁴, although the quantity is heavily dependent on the assumptions made about the clumpiness of the dust (if the dust is uniform, the observations imply a very low

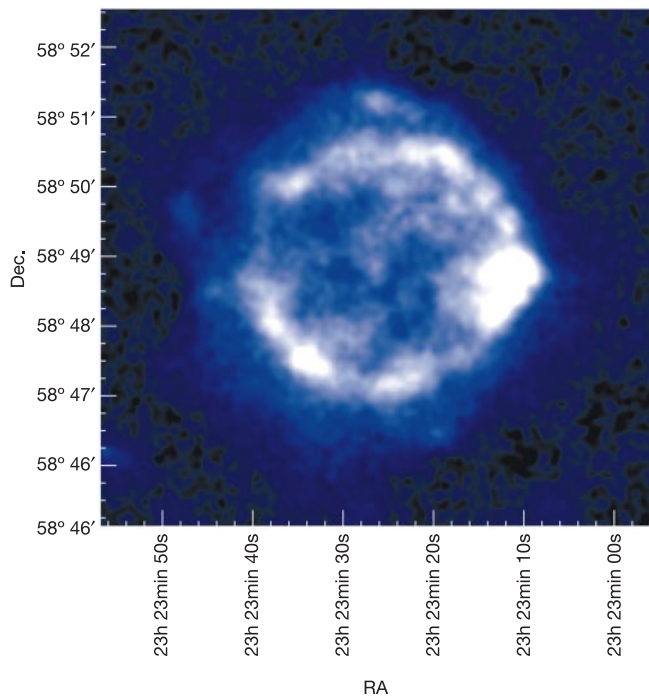


Figure 1 SCUBA $850\text{-}\mu\text{m}$ image of Cas A at a resolution of 15 arcsec . The data were taken in scan map mode and reduced in a standard way using SURF. Scan maps are known to produce artefacts on large angular scales on the final image and these remain the limitation on the accuracy of the absolute flux measurement and morphology. On the $850\text{-}\mu\text{m}$ image, the linear baseline removal left an area of diffuse emission to the west of the remnant which we believe was an artefact. After subtraction of a surface which removed this feature the loss of flux in the remnant was 5 Jy . We confirmed the accuracy of the absolute flux level of the maps by making new photometric measurements with SCUBA at several positions on the remnant in December 2002. The quoted 1σ uncertainty on the fluxes is a combination of calibration errors ($5\text{--}10\%$ and $15\text{--}20\%$ at 850 and $450\text{ }\mu\text{m}$), plus the range of fluxes from the maps which agreed with the photometry. The integrated fluxes were corrected for the error lobe contribution (derived from maps of Uranus) by dividing by 1.2 at $850\text{ }\mu\text{m}$ and 2.0 at $450\text{ }\mu\text{m}$.

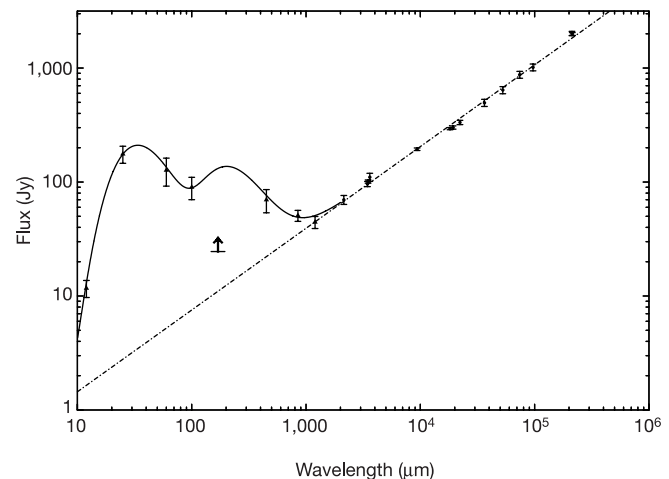


Figure 2 The SED of Cas A from the mid-infrared to the radio. All fluxes are integrated. The radio spectrum is fitted with a power-law slope with spectral index $\alpha = -0.72$. The radio data^{19,23–28} have been scaled to the epoch of our Cas A observation (1998.4) using a recent study²³ which found that the flux decrease of Cas A was independent of frequency at about $0.6\text{--}0.7\% \text{ yr}^{-1}$. The IRAS points are the averages of the literature values^{13,29} with the error bars indicating the range of fluxes. SCUBA fluxes are $50.8 \pm 5.6\text{ Jy}$ at $850\text{ }\mu\text{m}$ and $69.8 \pm 16.1\text{ Jy}$ at $450\text{ }\mu\text{m}$. The synchrotron contribution is 34.9 Jy at $850\text{ }\mu\text{m}$ and 22.1 Jy at $450\text{ }\mu\text{m}$. The fit to the far-infrared/submm points is for a two-temperature grey body with $\beta = 0.9$, $T_w = 112\text{ K}$, $T_c = 18\text{ K}$ and for 700 times more mass in the cold component than in the hot. We have fitted five parameters to six data points and therefore the fit is just constrained. The uncertainties in the SED parameters are given in Table 1. Clearly, the submm points at 450 and $850\text{ }\mu\text{m}$ lie above the extrapolation of the radio synchrotron spectrum. The point at 1.2 mm (ref. 28) is not of sufficient accuracy to determine an excess, but is consistent with our higher frequency measurements. Longer wavelength ISO measurements³⁰ of Cas A at $170\text{ }\mu\text{m}$ suggested that dust at $\sim 30\text{ K}$ may have been present (implying of order $0.15M_{\odot}$ of dust), however, the poor angular resolution of ISO at these wavelengths made a separation of the remnant and background emission impossible. The ISO flux is therefore a lower limit (R.J. Tufts, personal communication).

efficiency of condensation of heavy elements into dust grains ($\sim 0.001\%$), while if dust is very clumped this could rise to almost 100% in some species.) IRAS/ISO observations were not sensitive to cold dust at ≤ 25 K, and if the models are correct about the amount of dust produced in supernovae then the bulk of the dust must be cold. Such dust will emit most of its radiation at longer wavelengths and so the best place to search for cold dust emission is at sub-millimetre (submm) wavelengths (0.3–1 mm). Here, we present our analysis of submm data for Cas A, the youngest-known Galactic remnant. Cas A is the brightest radio source in the sky, and still produces significant synchrotron emission at submm wavelengths. Cas A is believed to be the remnant of the explosion of a massive ($20\text{--}30M_{\odot}$) progenitor star, which occurred around 320–340 yr ago at a distance of 3.4 kpc (refs 15, 16).

Cas A was observed with the SCUBA¹⁷ array on the James Clerk Maxwell Telescope (JCMT) in June 1998. The 850- μm image is presented in Fig. 1 and shows a ring-like morphology, similar to the X-ray and radio images^{18,19}. Over two-thirds of the 850- μm emission in the main ring is synchrotron produced by relativistic electrons spiralling in the intense magnetic field. Figure 2 presents the spectral energy distribution (SED) of Cas A from the radio to the mid-infrared, showing the clear excess due to dust at wavelengths of $\leq 850\mu\text{m}$. The synchrotron radiation displays a constant power-law behaviour over more than two decades in frequency with $S_{\nu} \propto \nu^{\alpha}$ where $\alpha = -0.72$, meaning that its contribution to the submm flux can be easily estimated and subtracted. The 450- μm image is shown in Fig. 3, the synchrotron contribution is now only one-third and emission from cold dust dominates, which is why the 450 and 850- μm images appear different. We can remove the synchrotron contribution from the submm maps by scaling an 83 GHz image¹⁹, which is the closest in frequency and resolution to our images. Figure 4 shows the result with contours denoting the 450- μm synchrotron subtracted flux.

The general similarity between the two wavelengths, once the synchrotron has been removed, is strong evidence that this is indeed emission from cold dust.

The rings on Fig. 4 indicate the position of the forward and reverse shocks as determined from Chandra X-ray data¹⁸. Most of the dust appears to be contained between the two, where the gas density as traced by the X-rays is greatest. We fitted a two-temperature grey-body SED to the synchrotron corrected infrared/submm fluxes (Fig. 2) determining temperatures of 112 K and 18 K for the two components. The SED parameters and their uncertainties are given in Table 1. The temperature of the hot dust is easily explained if it is co-extensive with the X-ray-emitting gas ($10^6\text{--}10^7$ K) and is heated by collisions with fast-moving electrons and ions¹³. The cold dust could be explained in a number of ways and a more detailed investigation will be presented elsewhere (L.D. *et al.*, manuscript in preparation). If the grains are very small ($<0.005\mu\text{m}$) they may cool to 15–20 K in the time between collisions with the gas particles. However, the dust SED might then be expected to show dust at all temperatures between 100–20 K, which is not apparent. Alternatively, the cold dust may be located in a very hot but diffuse phase of the gas, surrounding the bright, dense X-ray filaments. If the gas had a density of $<0.01\text{ cm}^{-3}$, then the grains could be this cold. The dust could also be contained in dense clumps which are not in pressure equilibrium with the more diffuse X-ray gas a model which has been suggested for SN1987A¹⁴. Finally, the cold dust could be in thermal equilibrium with the same X-ray gas that is responsible for heating the hot dust, if the cold grains are both very large (1–10 μm) and emit more efficiently than grains in the diffuse ISM. Evidence for unusual grain properties in Cas A is presented below.

The mass of dust can be estimated from the submm emission using $M_d = S_{\nu} D^2 / \kappa_{\nu} B(\nu, T)$, where D is the distance to the remnant

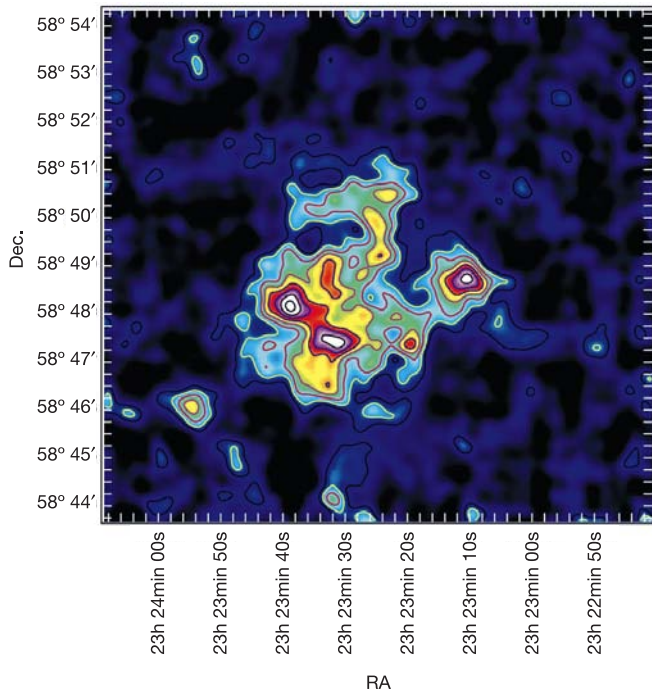


Figure 3 SCUBA 450- μm map, smoothed with a $21''$ gaussian. Colours and contours both represent the intensity of the 450- μm emission. Contours start at 2σ , with intervals of 1σ . The difference in morphology between this image and that at 850 μm (Fig. 1) is because emission from cold dust dominates at 450 μm , while two-thirds of the emission at 850 μm is synchrotron.

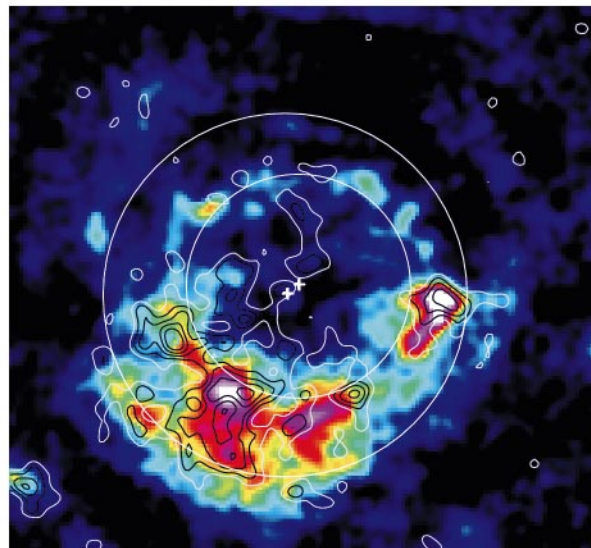


Figure 4 The 850 μm emission once the synchrotron has been subtracted using an 83-GHz image¹⁹. The box is 8.4 arcmin by 7.8 arcmin, north is up and east is left. Colours represent the 850 μm intensity, contours are the 450 μm emission with the synchrotron subtracted, starting at 3σ with increments of $+1\sigma$. The rings and crosses indicate the location (and centroids) of the forward and reverse shocks as determined from Chandra X-ray data¹⁸. The forward shock is at a mean radius of $153 \pm 12''$ and the reverse shock is at a mean radius of $95 \pm 10''$. The bulk of the dust emission appears to be bounded by the shocks, where the gas density is highest. There is a noticeable asymmetry in the dust emission.

Table 1 **Properties of Cas A**

SED parameters			Dust masses ($M_{\odot}$)		
β	T_{hot} (K)	T_{cold} (K)	κ_1	κ_2	κ_3
$0.9^{+0.8}_{-0.6}$	112^{+11}_{-21}	$18.0^{+2.8}_{-4.6}$	$M_{450} = 2.2^{+1.2}_{-0.2}$	$M_{450} = 3.8^{+2.0}_{-0.4}$	$M_{450} = 12.7^{+6.9}_{-1.1}$
			$M_{850} = 2.4^{+0.7}_{-0.3}$	$M_{850} = 6.2^{+1.7}_{-0.9}$	$M_{850} = 26.0^{+6.4}_{-4.1}$

SED parameters are the best least-squares fit to the data points. The errors were estimated using a bootstrap procedure in which the measured fluxes were perturbed according to their gaussian errors and re-fitted to find new parameters. The process was repeated 3,000 times and the uncertainties quoted are the 68% confidence interval derived from this method. Quoted dust masses are those from the best-fit SED parameters for each value of κ , and the ranges reflect the 68% confidence interval using the distribution produced by the bootstrap, essentially the uncertainty in dust mass due to the uncertainty in fluxes and SED fitting. We have used three values of κ which represent the average from numerous sources in the literature for different grain environments (to be presented in more detail elsewhere; L.D. *et al.*, manuscript in preparation.) (1) Laboratory measurements and theoretical modelling suggest that if grains are amorphous or have a clumpy, aggregate structure, κ can be very high: $0.6\text{--}1.1\text{ m}^2\text{ kg}^{-1}$ at $850\text{ }\mu\text{m}$. We take $\kappa_1(450) = 1.5\text{ m}^2\text{ kg}^{-1}$, $\kappa_1(850) = 0.76\text{ m}^2\text{ kg}^{-1}$. (2) Environments where dust may be newly formed, or modified by icy mantles also show high values: $0.16\text{--}0.8\text{ m}^2\text{ kg}^{-1}$. From observations of evolved stars, VRN (visual reflection nebulae), planetary nebulae and cold clouds in the Galaxy we take $\kappa_2(450) = 0.88\text{ m}^2\text{ kg}^{-1}$, $\kappa_2(850) = 0.3\text{ m}^2\text{ kg}^{-1}$. (3) Studies of extragalactic systems and diffuse ISM dust in the Galaxy support lower values of κ in the region of $0.04\text{--}0.15\text{ m}^2\text{ kg}^{-1}$ at $850\text{ }\mu\text{m}$. We take $\kappa_3(450) = 0.26\text{ m}^2\text{ kg}^{-1}$, $\kappa_3(850) = 0.073\text{ m}^2\text{ kg}^{-1}$.

(3.4 kpc), $B(\nu, T)$ is the Planck function at temperature T and κ is the dust mass absorption coefficient. The submm emission is clearly dominated by the cold dust, and therefore the cold temperature is the only one we need be concerned with. The value of κ is the main uncertainty as its value is not well determined and may vary with physical environment. Dust masses derived from the SED and different values of κ are in Table 1. Using ‘standard’ κ values appropriate to the diffuse ISM²⁰ (κ_3) gives dust masses which are uncomfortably high ($>7M_{\odot}$). Higher κ values, as measured in environments where dust may be newly formed, amorphous or coagulated (see Table 1) produce more sensible estimates (2–6 $M_{\odot}$), and thus the dust in Cas A appears to have different properties to that in the diffuse ISM. It is possible that grain processing in the diffuse ISM by ultraviolet photons and passage through dark clouds may alter the surface chemistry or shape of grains, causing a change in their emissivity. At an age of 320 yr, the dust in Cas A is relatively pristine and has not yet been subjected to the above events, which may contribute to its greater emissivity.

Was the dust produced in the supernova explosion, or could it be pre-existing dust which was swept up by the blast wave? If the remnant has swept up ISM material with a uniform density¹⁹ of $0.4\text{--}4\text{ H atoms per cm}^3$ then only $0.004\text{--}0.04M_{\odot}$ of dust should be present (for an ISM gas-to-dust ratio of 160). Higher ISM densities are possible, but in order for the mass of dust inferred in Cas A to have been swept up, densities of the order $200\text{--}400\text{ cm}^{-3}$ are required. This would mean Cas A has swept up $\sim 200\text{--}400M_{\odot}$ of material which is inconsistent with its dynamical state^{18,21}, believed to be just entering the Sedov stage (that is, it has swept up about as much mass as it ejected ($<15M_{\odot}$)). Another hypothesis for the bright ring seen in the X-ray and radio regions is that it is the interaction of the supernova with a stellar wind shell swept up by the Wolf–Rayet precursor^{16,22}. In this case, one could estimate that somewhere between $4\text{--}16M_{\odot}$ of material could have been lost in the red-supergiant phase and swept up into the ring. For this to be the source of the dust we observe requires an implausibly large quantity of freshly synthesized elements efficiently condensing into dust in the stellar wind material⁶ (which, unlike the supernova ejecta, is mainly H and He). We conclude that the most likely explanation for the large mass of dust in Cas A is that it was formed as a result of the supernova explosion.

Using our result, we estimate the current Galactic dust production rate from Type-II supernovae as $(7\text{--}18) \times 10^{-3}M_{\odot}\text{ yr}^{-1}$. Stellar wind⁶ sources produce $\sim 5 \times 10^{-3}M_{\odot}\text{ yr}^{-1}$, meaning that supernovae are dominant source of interstellar dust during most

of Galactic evolution. We have shown here that individual supernovae are capable of producing of the order of a solar mass of dust in the short times available in the early Universe; therefore the dust observed at high redshift is likely to have originated in supernovae. $\square$

Received 31 March; accepted 3 June 2003; doi:10.1038/nature01792.

- Bertoldi, F. *et al.* Dust emission from the most distant quasars. *Astron. Astrophys. Lett.* (in the press); also available at <http://www.arXiv.org/astro-ph/0305116> (2003).
- Archibald, E. N. *et al.* A submillimetre survey of the star formation history of radio galaxies. *Mon. Not. R. Astron. Soc.* **323**, 417–444 (2001).
- Hughes, D. H. *et al.* High-redshift star formation in the Hubble Deep Field revealed by a submillimetre-wavelength survey. *Nature* **394**, 241–247 (1998).
- Smail, I., Ivison, R. J. & Blain, A. W. A deep sub-millimetre survey of lensing clusters: A new window on galaxy formation and evolution. *Astrophys. J.* **490**, L5–L8 (1997).
- Dunne, L., Eales, S. A. & Edmunds, M. G. A census of metals at high and low redshifts and the connection between submillimetre sources and spheroid formation. *Mon. Not. R. Astron. Soc.* **341**, 589–598 (2003).
- Morgan, H. L. & Edmunds, M. G. Dust formation in early galaxies. *Mon. Not. R. Astron. Soc.* (in the press); also available at <http://www.arXiv.org/astro-ph/0302566> (2003).
- Todini, P. & Ferrara, A. Dust formation in primordial Type II supernovae. *Mon. Not. R. Astron. Soc.* **325**, 726–736 (2001).
- Clayton, D. D., Liu, W. & Dalgarno, A. Condensation of carbon in radioactive supernova gas. *Science* **283**, 1290–1292 (1999).
- Kozasa, T., Hasegawa, H. & Nomoto, K. Formation of dust grains in the ejecta of SN 1987A. II. *Astron. Astrophys.* **249**, 474–482 (1991).
- Woosley, S. E. & Weaver, T. A. The evolution and explosion of massive stars. II. Explosive hydrodynamics and nucleosynthesis. *Astrophys. J. Suppl. Ser.* **101**, 101–181 (1995).
- Lagage, P. O. *et al.* Dust formation in the Cassiopeia A supernova. *Astron. Astrophys.* **315**, L273–L276 (1996).
- Douvion, T., Lagage, P. O., Cesaesky, C. J. & Dwek, E. Dust in the Tycho, Kepler and Crab supernova remnants. *Astron. Astrophys.* **373**, 281–291 (2001).
- Dwek, E., Dinerstein, H. L., Gillet, F. C., Hauser, M. G. & Rice, W. L. Physical processes and infrared emission from the Cassiopeia A supernova remnant. *Astrophys. J.* **315**, 571–579 (1987).
- Lucy, L. B., Danziger, I. J., Gouiffes, C. & Bouchet, P. in *Supernovae: The 10th Santa Cruz Workshop in Astronomy and Astrophysics* (ed. Woosley, S. E.) 82 (Springer, New York, 1991).
- Reed, J. E., Hester, J. J., Fabian, A. C. & Winkler, P. F. The three-dimensional structure of the Cassiopeia A supernova remnant. I. The spherical shell. *Astrophys. J.* **440**, 706–721 (1995).
- García-Segura, G., Langer, N. & MacLow, M.-M. The hydrodynamic evolution of circumstellar gas around massive stars. *Astron. Astrophys.* **316**, 133–146 (1997).
- Holland, W. S. *et al.* SCUBA: a common-user submillimetre camera operating on the James Clerk Maxwell Telescope. *Mon. Not. R. Astron. Soc.* **303**, 659–672 (1999).
- Gottlieb, E. V. *et al.* CHANDRA detection of the forward and reverse shocks in Cassiopeia A. *Astrophys. J.* **552**, L39–L43 (2001).
- Wright, M., Dickel, J., Koralesky, B. & Rudnick, L. The supernova remnant Cassiopeia A at millimeter wavelengths. *Astrophys. J.* **518**, 284–297 (1999).
- Li, A. & Draine, B. T. Infrared emission from interstellar dust. II. The diffuse interstellar medium. *Astrophys. J.* **554**, 778–802 (2001).
- Agüeros, M. A. & Green, D. A. The bulk expansion of the supernova remnant Cassiopeia A at 151 MHz. *Mon. Not. R. Astron. Soc.* **305**, 957–965 (1999).
- Borkowski, K. J., Szmokowiak, A. E., Blondin, J. M. & Sarazin, C. L. A circumstellar shell model for the Cassiopeia A supernova remnant. *Astrophys. J.* **466**, 866–870 (1996).
- Reichart, D. E. & Stephens, A. W. The fading of supernova remnant Cassiopeia A from 38 MHz to 16.5 GHz from 1949 to 1999 with new observations at 1405 MHz. *Astrophys. J.* **537**, 904–908 (2000).
- Baars, J. W. M., Genzel, R., Pauliny-Toth, I. I. K. & Witzel, A. The absolute spectrum of CAS A—an accurate flux density scale and a set of secondary calibrators. *Astron. Astrophys.* **61**, 99–106 (1977).
- O’Sullivan, C. & Green, D. A. Constraints on the secular decrease in the flux density of CAS A at 13.5, 15.5 and 16.5 GHz. *Mon. Not. R. Astron. Soc.* **303**, 575–578 (1999).
- Mason, B. S. *et al.* An absolute flux density measurement of the supernova remnant Cassiopeia A at 32 GHz. *Astron. J.* **118**, 2908–2918 (1999).
- Liszt, H. & Lucas, R. 86 and 140 GHz radiocontinuum maps of the Cassiopeia A SNR. *Astron. Astrophys.* **347**, 258–265 (1999).
- Mezger, P. G., Tuffs, R. J., Chini, R., Kreysa, E. & Gemuend, H.-P. Maps of Cassiopeia A and the Crab Nebula at lambda 1.2 mm. *Astron. Astrophys.* **167**, 145–150 (1986).
- Saken, J. M., Fesen, R. A. & Shull, J. M. An IRAS survey of Galactic supernova remnants. *Astrophys. J. Suppl. Ser.* **81**, 715–745 (1992).
- Tuffs, R. J., *et al.* in *The Universe as Seen by ISO* (eds Cox, P. & Kessler, M. F.) 241–254 (ESA-SP 427, ESA Publications Division, ESTEC, Noordwijk, 1999).

Acknowledgements We thank M. Wright for providing us with the 83-GHz image, and W. Gear and D. Green for discussions. L.D. is supported by a PPARC postdoctoral fellowship and H.M. by a Cardiff University studentship. We are grateful to G. Davis for the use of Director’s time on the JCMT to obtain the new photometry data.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.D. (L.Dunne@astro.cf.ac.uk).

Efficient disruption of small asteroids by Earth's atmosphere

P. A. Bland* & N. A. Artemieva†

* Department of Earth Science and Engineering, Exhibition Road, Imperial College London, South Kensington Campus, London SW7 2AZ, UK

† Institute for Dynamics of Geospheres, Russian Academy of Sciences, Leninsky Prospect 38/6, Moscow, 117939 Russia

Accurate modelling of the interaction between the atmosphere and an incoming bolide is a complex task, but crucial to determining the fraction of small asteroids that actually hit the Earth's surface. Most semi-analytical approaches have simplified the problem by considering the impactor as a strengthless liquid-like object ('pancake' models^{1,2}), but recently a more realistic model has been developed that calculates motion, aerodynamic loading and ablation for each separate particle or fragment in a disrupted impactor^{3,4}. Here we report the results of a large number of simulations in which we use both models to develop a statistical picture of atmosphere-bolide interaction for iron and stony objects with initial diameters up to ~1 km. We show that the separated-fragments model predicts the total atmospheric disruption of much larger stony bodies than previously thought. In addition, our data set of >1,000 simulated impacts, combined with the known pre-atmospheric flux of asteroids with diameters less than 1 km⁵⁻¹², elucidates the flux of small bolides at the Earth's surface. We estimate that bodies >220 m in diameter will impact every 170,000 years.

We performed 16 simulations for stony impactors ('stones') using the separated fragments (SF) model, and 16 for iron impactors ('irons'), for bodies from 1 to 10⁸ kg, repeating each simulation for a given mass >20 times to derive average impact conditions (in total, >1,000 SF model simulations were performed). 'Pancake' model simulations were performed over the range 1–10¹² kg. When the model outputs are compared (Fig. 1), we find that 'pancake' and SF model estimates of total surviving material at the surface coincide tolerably well for irons, but the same is not true for stones. A 'pancake' model with spreading to twice the initial radius is typically chosen, which significantly overestimates impactor survivability for stones over the whole mass range (Fig. 1). SF and 'pancake' results only converge when we consider spreading to four times the initial radius (much larger than typically used) and only for initial masses >10⁷ kg. The 'pancake' model is not capable of providing a mass- or velocity-distribution for fragmented impactors, and therefore is not relevant to modelling production of terrestrial crater fields where the size of the largest crater is related to the largest surviving fragment. At larger masses (>10⁷ kg) this becomes less important, possibly as larger stones behave as a liquid-like 'swarm' of fragments.

The SF simulations allow us to define the various styles of meteoroid fragmentation exhibited by differing impactor types and masses. Meteoroids that are disrupted and dispersed in the atmosphere give rise to strewn fields at the surface. For lower-mass objects, the strewn field exists as meteorite fragments, and then (with increasing mass) meteorites plus impact pits; small, well separated craters with diameters ranging from metres to tens of metres; craters whose rims overlap; and finally, single craters produced by a swarm of poorly separated fragments. The masses that define the transition from one type of disrupted meteoroid impact to another are indicated in Fig. 1.

SF modelling also quantifies the very different survivability of iron and stony impactors. Over the mass range 10³–10⁷ kg, iron impactors transfer to the surface about three orders of magnitude more energy per unit area than stones: a fragmented iron impactor

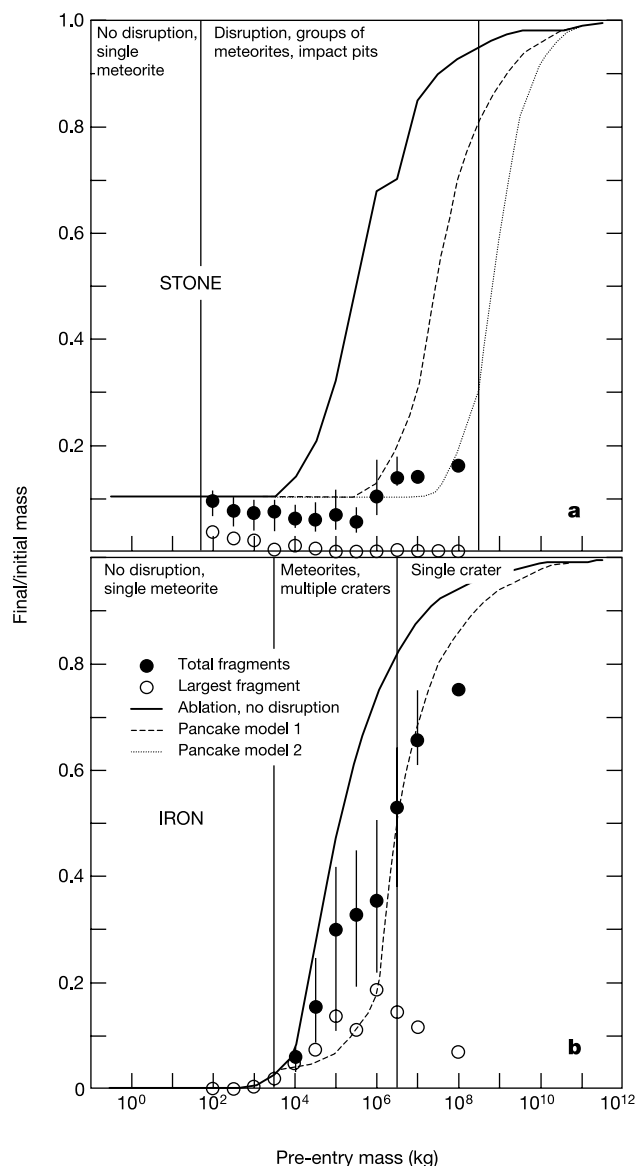


Figure 1 Results of SF and 'pancake' model simulations for stone and iron impactors. The SF model is described in detail elsewhere^{3,4}. The model takes into account successive fragmentation and ablation of individual fragments, and simulates the evolution of a meteoroid consisting of a variable number of solid fragments. Early attempts were made at modelling separated fragments^{23,24}, but most subsequent numerical approximations have taken the form of the more common 'pancake' model^{1,2}. While 'pancake' models treat the disrupted meteoroid as a deformable continuous liquid, the SF approximation allows us to define a mass- or velocity-distribution at the surface for fragments that create craters (high final velocity) or occur as meteorites (fragments with low final velocity). The production of crater fields by small fragmented asteroids may therefore be simulated. Two types of projectile are principally considered: irons with density of 7,800 kg m⁻³, ablation coefficient of 0.07 s² km⁻² and strength of 4.4 × 10⁸ dyn cm⁻² (for 1-kg samples)³, and stones with density of 3,400 kg m⁻³, ablation coefficient of 0.014 s² km⁻², and 10 × lower strength. The parameters for stones were chosen to define approximate upper limits on strength and density: larger stony bodies in the atmosphere, and carbonaceous bolides, may well have significantly lower strength and density. All simulations were at average asteroidal impact velocities and entry angles: 18 km s⁻¹ and 45°, respectively. The figure portrays the ratio of final mass (both the combined mass of all surviving fragments >100 g, and the largest single surviving fragment) to initial mass for stone (a) and iron (b) impactors. The 'pancake' model results are also shown: 'pancake model 1' is based on spreading to twice initial radius; 'pancake model 2', spreading to four times initial radius.

of 10^5 kg produces a similar crater field to a fragmented 10^8 kg stone. Even larger stony bodies of $\sim 10^8$ – 10^{10} kg are much less efficient at transferring energy to the surface than the equivalent iron impactor (Fig. 1). SF model simulations constrain the mass-, velocity- and size-distribution of those fragments, allowing us to derive morphologies of simulated crater fields. The SF results are in good agreement with terrestrial crater records, and also with available meteorite data. In meteorite falls, the proportion of stones decreases steadily at higher masses. Stony meteorites account for 96.6% of falls of <10 kg, falling to 87.1% at 50–100 kg, and 72.7% at 300–1,000 kg (extrapolating, falls $>4 \times 10^4$ kg will be $<5\%$ stones). Small terrestrial craters are also typically associated with iron impactors. Seventeen craters with diameters <1.5 km (impactors $\sim 2 \times 10^8$ kg) are associated with a given impactor type, and only

one is a stone¹³. In contrast, 18 craters >10 km have an inferred impactor type and 16 are stones¹³. This suggests that in the range $\sim 4 \times 10^4$ to 2×10^8 kg, $<5\%$ of terrestrial impactors are stones.

Our data set of simulations provides us with a template for understanding the atmosphere–bolide interaction for a range of pre-entry masses and impactor types. With a knowledge of the overall flux and the ratio of irons to stones (I/S) at the upper atmosphere, we can use these data to calculate the mass flux (both combined mass of all fragments >100 g, and the largest surviving fragment) at the Earth's surface.

The I/S ratio for the upper atmosphere can be inferred from near-Earth object (NEO) and main-belt asteroid spectroscopy, meteorite composition, and impactor types in large terrestrial craters. Assuming that iron asteroids are within the X-class, spectroscopy places

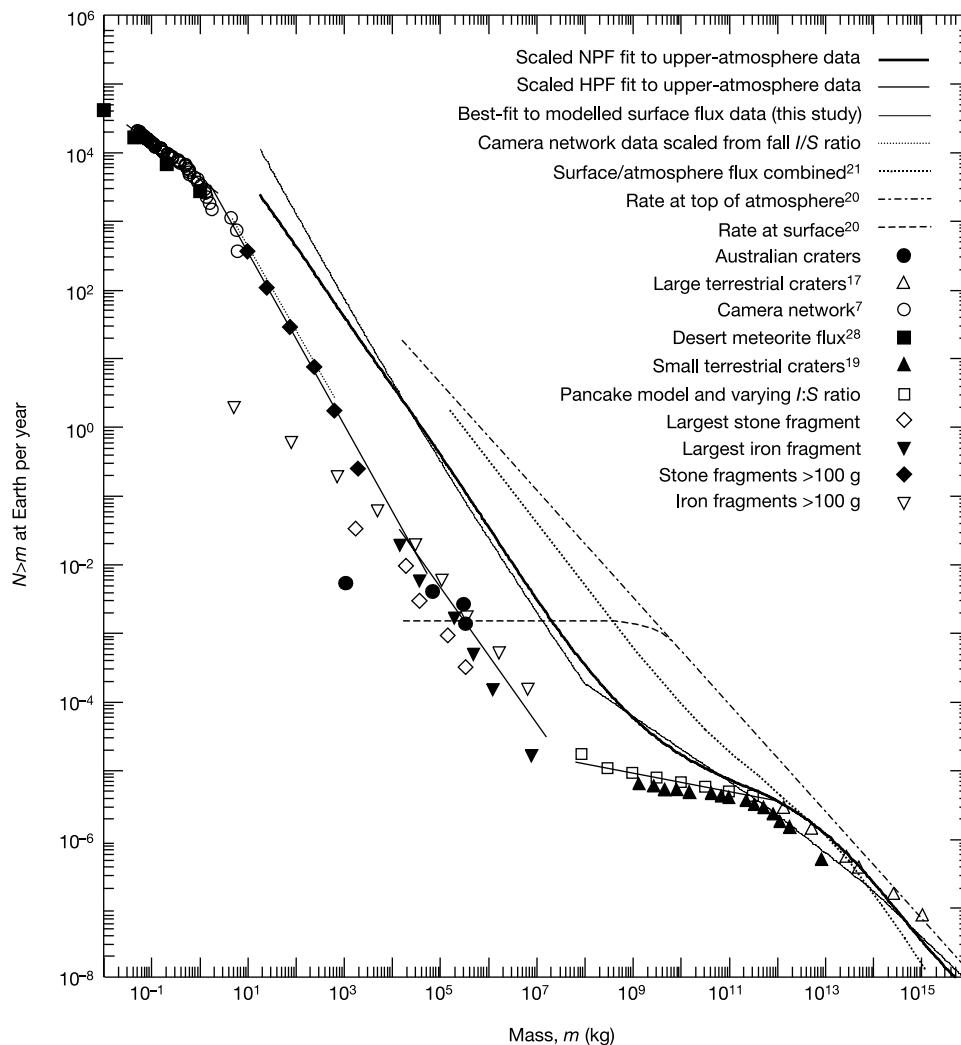


Figure 2 Estimates of impact rate for both the top of the atmosphere and the Earth's surface. Impactor flux is expressed in terms of number of events N greater than mass m per year at Earth (pre-entry mass for the upper-atmosphere data, final mass for surface flux estimates). Crater data are scaled to projectile diameter^{25,26}, and the Hartmann and Neukum production functions (HPF and NPF) derived from the lunar mare crater record^{18,27} scaled to fit the upper-atmosphere data^{5–12,17}. Earlier impact rate estimates are shown for reference^{20–22}. Some combine surface and atmosphere flux estimates²¹, while others define separate estimates²⁰. We scale SF model results for 10 – 10^3 kg bodies based on the NPF for the upper atmosphere, taking into account the decreasing proportion of stone meteorite falls of 1 – 10^3 kg. We also derive an estimate by extrapolating from fireball data for small meteorites on the ground⁵ to larger bodies based on the observed depletion in stone meteorite falls. The SF data and this curve are

coincident. Over the mass range 4×10^4 – 1×10^8 kg, SF model data constrain the fraction of the impactor that reaches the surface with high energy. A surface flux is estimated by scaling the upper-atmosphere flux given the ratio of irons to stones (I/S) at the upper atmosphere and the calculated final mass derived from SF simulations. Modelling separated fragments becomes unwieldy for objects $>10^8$ kg. However, SF and 'pancake' model ($\times 4$ spreading) estimates of final mass for initial bodies $>10^7$ kg converge (Fig. 1): our data indicate these objects occur as a closely separated liquid-like 'swarm' of fragments at the surface. We therefore scale the NPF for the upper atmosphere to a surface flux using 'pancake' model results for objects $>10^8$ kg, and a varying I/S ratio based on the depletion in craters formed by stony impactors over the crater diameter interval 1.5 – 20 km.

upper limits on the abundance of irons among main-belt asteroids¹⁴ (~8%) and NEOs¹⁵ (~10%). If we only consider M-class asteroids (a sub-type within the X-class), the proportions are ~5% of main-belt asteroids¹⁶ and ~2% of NEOs¹⁵. However, it should be noted that a number of M-class asteroids show absorption features in their infrared spectra which indicate that they are not metallic, so the real number of iron asteroids may be lower. For meteorites <10 kg, ~3% are irons, and of the craters >10 km in diameter where the impactor type is known, ~10% were formed by irons. Taking these data together, to a first approximation, we estimate 5% irons (within a factor of two) at the upper atmosphere.

Over recent years, the impactor flux for the upper atmosphere has been well constrained over specific portions of the mass distribution using various techniques^{5–12,17}. Ivanov¹⁸ showed that the NEO and main-belt asteroid mass distribution^{9–12}, and upper-atmosphere meteoroid data⁶, matches the cumulative size–frequency distribution of lunar mare craters (the Neukum or Hartmann production functions—NPF or HPF). NPF and HPF curves are used in crater counting studies to date surfaces of all ages throughout the inner Solar System. Since the overall shape of the curve is likely to apply to the current impactor population at 1 AU, we fit the upper-atmosphere data with a scaled NPF and HPF (converted to projectile diameter). Over a mass range of 1–10¹⁶ kg the fit is a good one—camera network data⁵, recent US Department of Defense satellite data⁷, large NEO data^{9–11} and the data set of large terrestrial craters¹⁷ all fall within a factor of three of this curve (some earlier satellite data^{6,8} and modelled small NEO data¹² lie off this curve).

We now scale the NPF projectile curve for the upper-atmosphere flux to an impact rate at the Earth's surface using data from our simulations (Fig. 2). Previous estimates for the flux at the surface and the top of the atmosphere, and the scaled NPF and HPF curves for the upper atmosphere, are shown for comparison. As noted, the terrestrial crater record is incomplete over the mass range of interest. However, if we select recent (<0.05 Myr ago) small (<1 km) impacts in Australia, an area where arid conditions limit erosion and obscuring vegetation is at a minimum, we observe a close match to our modelled flux estimate for bodies of ~10⁵ kg (Fig. 2). For objects <10⁵ kg the Australia data depart from our modelled curve, but even minimal erosion would remove craters <100 m in diameter. For bodies of 10⁹–10¹² kg, our modelled curve is essentially identical to the terrestrial crater data¹⁹. The implication is that severe atmospheric disruption of stony impactors is the cause of the departure from a primary NEO-derived flux for 1.5–20-km craters on Earth.

Our simulations predict atmospheric fragmentation for much larger objects than previously thought. For stony asteroids with diameters up to several hundred metres, this means that the bulk of the energy is deposited in the atmosphere, rather than at the surface, but although similarly sized irons produce single craters, fragmentation is still an issue. The largest body we have attempted to model with the SF numerical approximation is an iron impactor with a pre-entry mass of 10¹⁰ kg. Even for a comparatively large high-strength object such as this, we find that aerodynamic loading is sufficient to disrupt the impactor (although the fragments will hit together in a confined area, and form a single complex crater of similar diameter to that which would result from the impact of a coherent projectile).

In terms of defining impact rates, it is useful to separate Tunguska-like events (typically, stony objects that penetrate to the lower atmosphere before massive disruption occurs), and cratering events. Although only the latter will have a lasting surface expression, both have significance for defining the hazard to human populations. Stony bodies of >10⁸ kg can penetrate low enough in the atmosphere before disruption to affect the surface, so the NPF/HPF fit to the upper-atmosphere data for bodies larger than 10⁸ kg defines the approximate rate at which Tunguska-like events will occur. The

significance of the curvature in the NPF/HPF, and its effect on impact rate estimates at Earth, does not appear to have been recognized in some earlier studies²⁰. Over much of the mass distribution, the NPF/HPF fit to the recent upper-atmosphere data suggest a lower flux than previous estimates^{20–22} (Fig. 2).

For the flux at the Earth's surface, our data suggest that impactors with largest fragment 3–5 m, capable of forming ~100-m craters, will strike the Earth every 200–400 years (>95% will be irons), larger 50-m-diameter impactors occur every 9.7×10^4 years (~90% irons), and 220-m impactors occur every 1.7×10^5 years (~50% irons). As we chose conservatively high values for stony bolide strength and density, and as it is entirely possible that the proportion of irons at the top of the atmosphere is <5%, our modelled flux curve probably represents an upper limit on surface impact rate for the stated flux at the upper atmosphere.

In terms of the impact hazard, these data have specific relevance for quantifying the threat posed by impact-generated tsunamis. It has been shown that the peak contributors to potentially hazardous tsunamis (those >5 m in height) are impactors of ~220-m diameter²⁰. The same study derived an estimate for the frequency of impacts of 220-m bolides of 1 every 3,000–4,000 years (ref. 20). Our data suggest an impact rate ~50 × lower. Although airburst events for stony bolides of ~200-m diameter may occur somewhat more frequently than surface impacts (~1 every 50,000–60,000 years), the rate is still low, and such events are unlikely to couple their energy efficiently enough into the ocean surface to generate large tsunamis. The implication is that the hazard posed by impact-generated tsunamis is lower than previously thought. This analysis is significant in redefining the asteroid impact hazard, and proposals to extend the survey of NEOs down to much smaller objects may need to have their search strategies reviewed in light of it. □

Received 30 January; accepted 21 May 2003; doi:10.1038/nature01757.

1. Chyba, C. F., Thomas, P. J. & Zahnle, K. J. The 1908 Tunguska explosion: Atmospheric disruption of a stony asteroid. *Nature* **361**, 40–44 (1993).
2. Hills, J. G. & Goda, M. P. The fragmentation of small asteroids in the atmosphere. *Astron. J.* **105**, 1114–1144 (1993).
3. Artemieva, N. A. & Shuvalov, V. V. Motion of a fragmented meteoroid through the planetary atmosphere. *J. Geophys. Res.* **106**, 3297–3310 (2001).
4. Shuvalov, V. V., Artemieva, N. A. & Trubetskaya, I. A. Simulating the motion of a disrupted meteoroid with allowance for evaporation. *Sol. Syst. Res.* **34**, 49–60 (2000).
5. Halliday, I., Griffin, A. A. & Blackwell, A. T. Detailed data for 259 fireballs from the Canadian camera network and inferences concerning the influx of large meteoroids. *Meteorit. Planet. Sci.* **31**, 185–217 (1996).
6. Nemtchinov, I. V. *et al.* Assessment of kinetic energy of meteoroids detected by satellite-based light sensors. *Icarus* **130**, 259–274 (1997).
7. Brown, P., Spalding, R. E., ReVelle, D. O., Tagliaferri, E. & Worden, S. P. The flux of small near-Earth objects colliding with the Earth. *Nature* **420**, 294–296 (2002).
8. ReVelle, D. O. Historical detection of atmospheric impacts by large bolides using acoustic-gravity waves. *Ann. NY Acad. Sci. USA* **822**, 284–302 (1997).
9. Morbidelli, A., Jedicke, R., Bottke, W. F., Michel, P. & Tedesco, E. F. From magnitudes to diameters: The albedo distribution of near-Earth objects and the Earth collision hazard. *Icarus* **158**, 329–343 (2002).
10. Rabinowitz, D., Helin, E., Lawrence, K. & Pravdo, S. A reduced estimate of the number of kilometre-sized near-Earth asteroids. *Nature* **403**, 165–166 (2000).
11. Stuart, J. S. A near-Earth asteroid population estimate from the LINEAR survey. *Science* **294**, 1691–1693 (2001).
12. Harris, A. W. in *Proc. Asteroids, Comets, Meteors 2002* (European Space Agency, Berlin, in press).
13. Earth Impact Database (<http://www.unb.ca/passc/impactdatabase/>) (accessed 20 November 2002).
14. Bus, S. J. & Binzel, R. P. Phase II of the small main-belt asteroid spectroscopic survey: A feature-based taxonomy. *Icarus* **158**, 146–177 (2002).
15. Binzel, R. P., Lupishko, D. F., Di Martino, M., Whiteley, R. J. & Hahn, G. J. in *Asteroids III* (eds Bottke, W. F., Cellino, A., Paolicchi, P. & Binzel, R. P.) 255–271 (Univ. Arizona Press, Tucson, 2003).
16. Tholen, D. J. in *Asteroids II* (eds Binzel, R. P., Gehrels, T. & Matthews, M. S.) 1139–1150 (Univ. Arizona Press, Tucson, 1989).
17. Grieve, R. A. F. & Shoemaker, E. M. in *Hazards due to Comets and Asteroids* (ed. Gehrels, T.) 417–462 (Univ. Arizona Press, Tucson, 1994).
18. Ivanov, B. A., Neukum, G., Bottke, W. F. & Hartmann, W. K. in *Asteroids III* (eds Bottke, W. F., Cellino, A., Paolicchi, P. & Binzel, R. P.) 89–101 (Univ. Arizona Press, Tucson, 2003).
19. Hughes, D. W. A new approach to the calculation of the cratering rate of the Earth over the last 125 ± 20 Myr. *Mon. Not. R. Astron. Soc.* **317**, 429–437 (2000).
20. Ward, S. N. & Asphaug, E. Asteroid impact tsunami: A probabilistic hazard assessment. *Icarus* **145**, 64–78 (2000).
21. Chapman, C. R. & Morrison, D. Impacts on the Earth by asteroids and comets: Assessing the hazard. *Nature* **367**, 33–39 (1994).
22. Morrison, D., Chapman, C. R. & Slovic, P. in *Hazards due to Comets and Asteroids* (ed. Gehrels, T.) 59–91 (Univ. Arizona Press, Tucson, 1994).

23. Passey, Q. R. & Melosh, H. J. Effects of atmospheric breakup on crater field formation. *Icarus* **42**, 211–233 (1980).
24. Melosh, H. J. *Impact Cratering: A Geological Process* (Oxford Univ. Press, Oxford, 1989).
25. Schmidt, R. M. & Housen, K. R. Some recent advances in the scaling of impact and explosion cratering. *Int. J. Impact Eng.* **5**, 543–560 (1987).
26. Ivanov, B. A., et al. in *Chronology and Evolution of Mars* (ed. Kallenbach, R.) 87–104 (Kluwer, Dordrecht, 2001).
27. Hartmann, W. K. Martian cratering VI: Crater count isochrons and evidence for recent volcanism from Mars Global Surveyor. *Meteorit. Planet. Sci.* **34**, 167–177 (1999).
28. Bland, P. A. et al. The flux of meteorites to the Earth over the last 50,000 years. *Mon. Not. R. Astron. Soc.* **283**, 551–565 (1996).

Acknowledgements We thank B. Ivanov, W. Hartmann and P. Brown for providing cratering data, flux data, and for discussions, and B. Ivanov, H. J. Melosh, V. Shuvalov, J. Morgan, E. Pierazzo and M. Gounelle for suggestions that improved earlier drafts of this manuscript. This work benefited greatly from comments and suggestions from C. Chapman. N.A. thanks RFBR for support, and P.A.B. thanks the Royal Society for support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.A.B. (p.a.bland@imperial.ac.uk).

Nanometre-scale displacement sensing using a single electron transistor

Robert G. Knobel*† & Andrew N. Cleland*

* Department of Physics and iQUEST, University of California, Santa Barbara, California 93106, USA

It has been a long-standing goal to detect the effects of quantum mechanics on a macroscopic mechanical oscillator^{1–3}. Position measurements of an oscillator are ultimately limited by quantum mechanics, where ‘zero-point motion’ fluctuations in the quantum ground state combine with the uncertainty relation to yield a lower limit on the measured average displacement. Development of a position transducer, integrated with a mechanical resonator, that can approach this limit could have important applications in the detection of very weak forces, for example in magnetic resonance force microscopy⁴ and a variety of other precision experiments^{5–7}. One implementation that might allow near quantum-limited sensitivity is to use a single electron transistor (SET) as a displacement sensor^{8–11}: the exquisite charge sensitivity of the SET at cryogenic temperatures is exploited to measure motion by capacitively coupling it to the mechanical resonator. Here we present the experimental realization of such a device, yielding an unequalled displacement sensitivity of $2 \times 10^{-15} \text{ m Hz}^{-1/2}$ for a 116-MHz mechanical oscillator at a temperature of 30 mK—a sensitivity roughly a factor of 100 larger than the quantum limit for this oscillator.

A classical simple harmonic oscillator, in equilibrium with its environment at temperature T , will have an average total energy $k_B T$. The position of the oscillator fluctuates continuously, with a root mean square displacement amplitude $\delta x = (k_B T / m \omega_0^2)^{1/2}$ for an oscillator of mass m and resonant frequency $f_0 = \omega_0 / 2\pi$. This classical displacement amplitude can be made arbitrarily small by reducing the temperature. One implication of quantum mechanics, however, is that the quantized nature of the oscillator energy yields an intrinsic fluctuation amplitude, the ‘zero-point motion’ $\delta x_{zp} = (\hbar / 2m\omega_0)^{1/2}$. This is achieved for temperatures T well below the energy quantum, $T \ll T_Q \equiv \hbar\omega_0 / k_B$. A second implication

of quantum mechanics is that the instrument used to measure the position of the oscillator will necessarily perturb it, further limiting the possible measurement resolution, as quantified by the Heisenberg uncertainty principle. A continuous measurement of the average position of an oscillator, using a quantum-limited amplifier, is thus limited to $\sqrt{2}$ times the zero-point motion, or $\delta x_{\text{meas}} \approx (\hbar / m\omega_0)^{1/2}$. We note that ‘back-action evading’ techniques have in principle unlimited measurement precision, although these yield less information and are still subject to zero-point fluctuations if the measurement is slower than the oscillator relaxation time^{1,12}.

A 1-GHz nanomechanical flexural resonator was recently demonstrated¹³. This resonator would have $T_Q = 50 \text{ mK}$, so when operated on a dilution refrigerator at 10 mK, the inequality $T \ll T_Q$ could be reached. A resonator with similar dimensions would be a candidate for detecting the transition from classical to quantum noise, because the small mass gives a relatively large zero-point displacement noise, $\delta x_{zp} \approx 2 \times 10^{-14} \text{ m}$. In terms of the spectral displacement noise density, this corresponds to $S_x^{1/2} = 5 \times 10^{-18} \text{ m Hz}^{-1/2}$ on resonance, fairly small due to the relatively low quality factor $Q \approx 500$. Techniques to measure the displacement of large-scale resonators, such as optical interferometry or electrical parametric transducers^{1,14,15}, can achieve better displacement noise limits than this, but the larger mass of the resonator makes reaching the quantum limit more difficult. Such techniques do not scale well to nanomechanical resonators, and other techniques more applicable to these size scales^{16,17} are unlikely to approach quantum-limited sensitivity. The single electron transistor provides a possible system in which sufficient sensitivity can be achieved.

The single-electron transistor (SET) consists of a conducting island separated from leads by low-capacitance, high-resistance tunnel junctions. The current through the SET is modulated by the charge induced on its gate electrode, with a period e , the charge of one electron. The SET is the most sensitive electrometer^{18,19}, with a demonstrated sensitivity below $10^{-5} e \text{ Hz}^{-1/2}$. The motion of a nanomechanical resonator may be detected by capacitively coupling the gate of the SET to a metal electrode placed on the resonator, and biasing the electrode at a constant voltage V_{beam} (see Fig. 1). The capacitance C between the SET and the beam then has a coupled charge $q = V_{\text{beam}} C$. As the beam vibrates in the x direction, in the

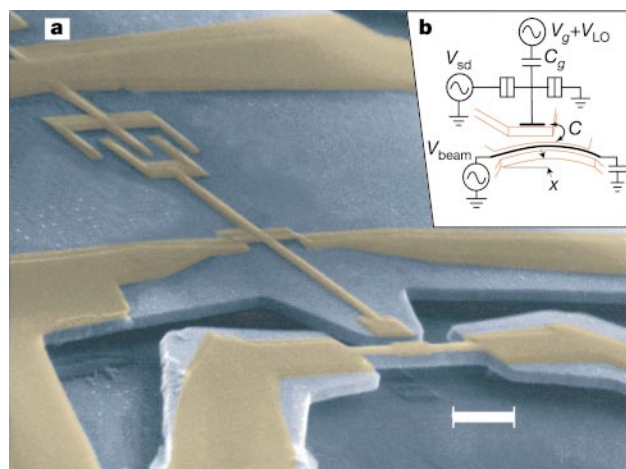


Figure 1 The device used in the experiment. **a**, Scanning electron micrograph of the device, showing the doubly clamped GaAs beam, and the aluminum electrodes (coloured) forming the single electron transistor and beam electrode. Scale bar, 1 μm. The Al/AIO_x/Al tunnel junctions have approximately 50 × 50 nm² overlap. **b**, A schematic of the mechanical and electrical operation of the device.

† Present address: Department of Physics, Queen's University, Kingston, Ontario K7L 3N6, Canada.

plane of the device, the resulting variation in capacitance ΔC will modulate the charge induced on the SET, $\Delta q = V_{\text{beam}} \Delta C$, changing the SET source-drain current. As the voltage V_{beam} is increased, the charge modulation Δq and the sensitivity to the resonator motion will increase. However, the source-drain current is due to the stochastic flow of electrons through the SET, so the centre island's voltage fluctuates randomly. This causes a fluctuating 'back-action' force on the beam. This force increases as V_{beam} increases, resulting in a voltage for which the total noise is minimized. The displacement sensitivity at this optimal voltage is calculated^{9,10} to be of order $10^{-16} \text{ m Hz}^{-1/2}$, approaching the sensitivity needed to measure quantum effects.

Our device (Fig. 1) consists of a $3 \mu\text{m}$ long $\times$ 250 nm wide $\times$ 200 nm thick doubly-clamped beam of single-crystal GaAs, capacitively coupled to an aluminum SET, located 250 nm from the beam. The beam was patterned using electron beam lithography and etched from a GaAs heterostructure using a sequence of reactive ion etching and dilute HF wet etching²⁰. The SET was formed through double-angle shadow evaporation, using a pattern defined in a second step of electron-beam lithography^{21,22}. The device was mounted on a dilution refrigerator and cooled to 30 mK . All electrical leads were filtered²³. An out-of-plane 8 T magnetic field was applied to drive the aluminum out of the superconducting state, and to provide a field for actuation and sensing of the beam motion.

The beam was driven using the magnetomotive technique, in which an alternating current I through the beam electrode, along its length L in the presence of the magnetic field B , generates a Lorentz force $F = ILB$. We measured the induced electromotive force (EMF) $\mathcal{E}$ developed along the beam owing to its resulting motion through the magnetic field^{16,24,25} by measuring the reflected and transmitted power. The beam vibrates in its fundamental in-plane mode with a resonant frequency $f_0 = 116.7 \text{ MHz}$ (see Fig. 2). The measured EMF fits well to a lorentzian function with a quality factor $Q \approx 1700$. The beam displacement Δx is related to the EMF $\mathcal{E}$ by²⁵:

$$\Delta x = \frac{\mathcal{E}}{\xi LB \omega} \quad (1)$$

where $\xi = 0.523$ is an integration constant. We thus determine the displacement Δx as a function of power P , as shown in Fig. 2,

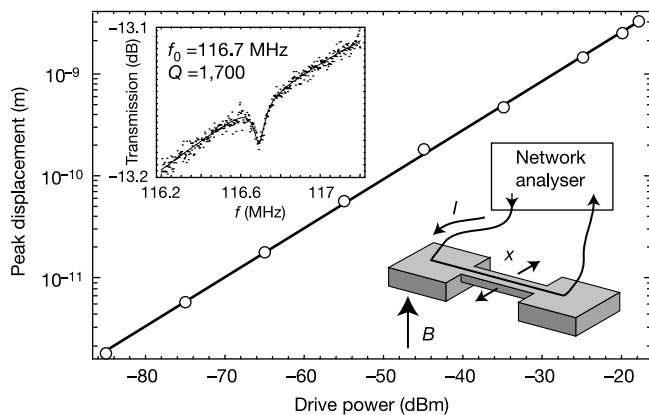


Figure 2 Magnetomotive measurements of the beam resonance. Upper inset, Raw magnetomotive transmission data (20 averages, 10 Hz bandwidth) as a function of drive frequency f , for a drive power $P = -85 \text{ dBm}$, with a fit of a Lorentzian line shape with a linear background. The fit yields the resonance frequency $f_0 = 116.7 \text{ MHz}$ and quality factor $Q = 1700$. The main plot shows the measured midpoint displacement Δx of the beam driven at its frequency resonance $f = f_0$, as a function of magnetomotive drive power P at 30 mK and in an 8 T magnetic field. The fit response $\Delta x = 839\sqrt{P} \text{ nm}/\text{W}^{1/2}$ (solid line), with the beam's known physical geometry (lower inset), corresponds to an effective spring constant $k = 1,530 \text{ N m}^{-1}$ and resonator mass $m = 2.84 \times 10^{-15} \text{ kg}$.

yielding the resonator mass $m = 2.84 \times 10^{-15} \text{ kg}$ and effective spring constant $k = 1,530 \text{ N m}^{-1}$.

The SET was characterized using low-frequency electrical measurements, giving a series junction resistance $R_{J1} + R_{J2} = 200 \text{ k}\Omega$, a gate capacitance $C_g = 0.59 \text{ fF}$, a coupling capacitance between the beam electrode and the SET of $C = 0.13 \text{ fF}$, and a total SET island capacitance $C_{\Sigma} = 1.3 \text{ fF}$. In order to detect motion at the resonant frequency of the beam, the SET was operated as a mixer¹¹, with a local oscillator (LO) voltage V_{LO} applied to the gate at a frequency offset from the drive frequency by an intermediate frequency (IF) of less than 1 kHz ($f_{\text{IF}} = |f_{\text{LO}} - f|$). The SET was biased so as to give optimum mixing signal, with $V_{\text{sd}} \approx e/C_{\Sigma}$ and the dc gate voltage adjusted to give peak dc current. The V_{LO} amplitude was set to give maximum signal gain, with coupled charge at the LO frequency $q_{\text{LO}} \approx 0.4e$. The voltage across the SET was measured at the intermediate frequency.

Figure 3a shows the mixer signal for a drive power $P = -125 \text{ dBm}$. Using the magnetomotive calibration, this corresponds to an extrapolated amplitude of $\Delta x = 2.3 \times 10^{-14} \text{ m}$ on resonance. The charge Δq detected by the SET is due to the change in coupling capacitance times the beam voltage, $\Delta q = V_{\text{beam}} \Delta C$, as well as secondary signals due to the induced EMF, and capacitively coupled charge due to cable resonances and ohmic losses. The former is very small compared to the direct signal (less than 1%), and the second varies slowly with frequency. Thus the total signal detected by the SET has the form of a Lorentzian, plus a background with an arbitrary phase, as shown in Fig. 3a.

The quality factor Q and resonant frequency f_0 obtained using the SET were the same as those from the magnetomotive technique, and the amplitude varied as $P^{1/2}$. The SET response is found to be 9.9

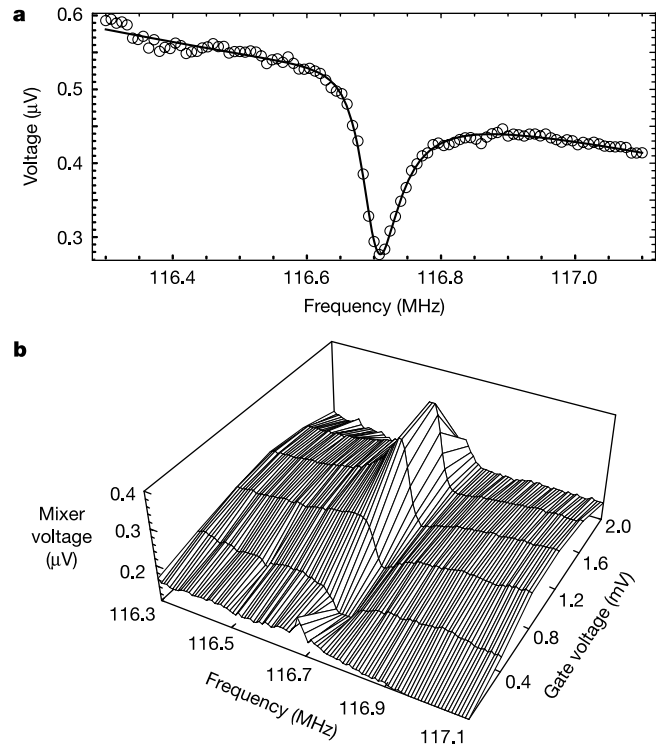


Figure 3 Single electron transistor measurement of the beam motion. **a**, Intermediate frequency voltage (SET mixer output) as a function of beam drive frequency, for $P = -125 \text{ dBm}$ drive power. The solid line is a fit to a Lorentzian plus linear background with arbitrary phase. The beam electrode had a d.c. voltage $V_{\text{beam}} = 5 \text{ V}$, the LO power was -70 dBm , and the IF was 151 Hz . The SET gate voltage was tuned to give maximum signal. **b**, IF voltage for $P = -115 \text{ dBm}$ drive power, as a function of d.c. gate voltage and drive frequency. The slowly varying background interferes with the resonance signal, whose phase varies with the gate voltage.

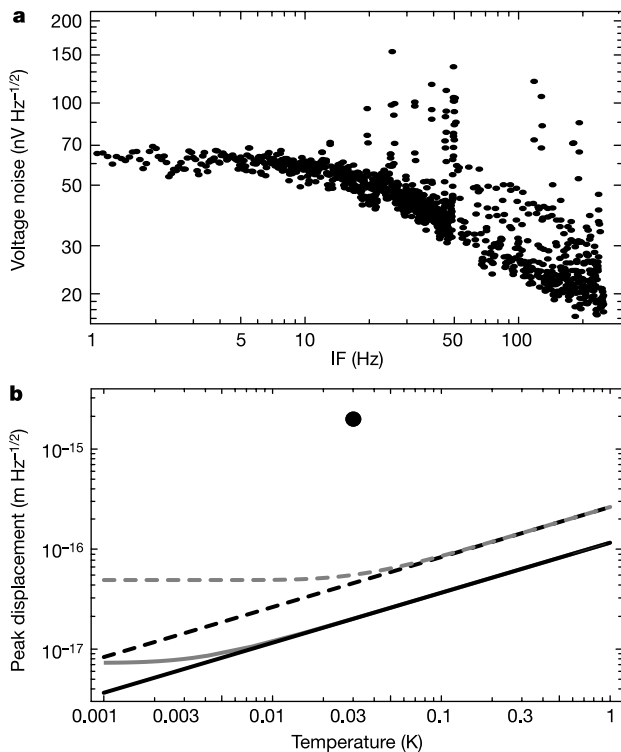


Figure 4 Noise and quantum limits for the device. **a**, Output voltage noise $S_V^{1/2}$ as a function of mixer intermediate frequency in the same configuration as used for the driven beam measurements ($L_O = -70$ dBm, $f_{LO} = 100$ MHz, $V_{sd} \approx e/C_S$, $V_{beam} = 5$ V, $B = 8$ T, $T = 30$ mK). For the displacement measurements, an intermediate frequency of 151 Hz was used, corresponding to a noise level of about $20 \text{ nV Hz}^{-1/2}$. **b**, Calculated displacement noise for two resonators, the resonator measured in this experiment (solid lines) and for a hypothetical 1-GHz resonator (dashed lines), fabricated from GaAs with dimensions $450 \times 50 \times 40$ nm and $Q = 10^4$. The black lines correspond to the thermal noise and the grey lines include the back-action noise and zero-point motion in an ideal quantum-limited position measurement. The measured noise of our device is indicated by the black circle.

$\Delta x \sqrt{V_{\mu\text{m}}^{-1}}$, proportional to the displacement Δx , when the SET is biased optimally, with a LO frequency offset from the beam drive signal by 151 Hz. As the SET gate voltage is varied from the optimum value, Fig. 3b, the sensitivity decreases and interference with the background becomes more pronounced. The resonant signal decreases linearly with magnetic field B , and increases linearly with V_{beam} .

Figure 4a shows the output voltage noise of the SET mixer. For the intermediate frequency used (151 Hz), the noise ($\sim 20 \text{ nV Hz}^{-1/2}$) is dominated by the second-stage amplifier. The displacement sensitivity can be calculated from the SET response: We find a peak sensitivity of $S_x^{1/2} \approx 2.0 \times 10^{-15} \text{ m Hz}^{-1/2}$.

We can now evaluate the potential for quantum-limited detection: In Fig. 4b we show the calculated thermal and quantum noise for this beam, and for a 1-GHz resonator, which would be in its quantum ground state at the base temperature of our dilution refrigerator. We see that the detector noise is ~ 100 times larger than the thermal noise of the measured beam, and about a factor of ~ 20 larger than the quantum noise of a 1-GHz resonator with $Q = 10^4$. By using improved electronics and impedance matching, the second-stage noise could be reduced by roughly a factor of ten, and better lithography could reduce the SET-beam separation, increasing the SET response. A higher quality factor would yield larger signals in both the thermal and quantum limits; experimentally the Q is observed to reduce as the resonators are made smaller, at least partly due to surface damage and contamination. Flash

heating of resonators in vacuum has been demonstrated to yield significantly higher quality factors^{26,27}, a technique which could be applied here. The expected increase in displacement sensitivity using a combination of these techniques could, in a second-generation device, allow measurement in the quantum-limited regime.

In conclusion, we have demonstrated an ultrasensitive, potentially quantum-limited displacement sensor based on a single-electron transistor, enabling us to read out the motion of a nanomechanical resonant beam at its resonant frequency. The device has a displacement sensitivity of $2.0 \times 10^{-15} \text{ m Hz}^{-1/2}$ at the 116.7-MHz resonant frequency of the mechanical beam, limited by the noise in the conventional electronics. □

Received 28 March; accepted 29 May 2003; doi:10.1038/nature01773.

1. Bocko, M. F. & Onofrio, R. On the measurement of a weak classical force coupled to a harmonic oscillator: Experimental progress. *Rev. Mod. Phys.* **68**, 755–790 (1996).
2. Braginsky, V. B. & Khalili, F. Y. *Quantum Measurement* (Cambridge Univ. Press, Cambridge, 1992).
3. Cho, A. Researchers race to put the quantum in mechanics. *Science* **299**, 36–37 (2002).
4. Sidles, J. A. *et al.* Magnetic resonance force microscopy. *Rev. Mod. Phys.* **67**, 249–265 (1995).
5. Armour, A. D., Blencowe, M. P. & Schwab, K. C. Entanglement and decoherence of a micromechanical resonator via coupling to a cooper-pair box. *Phys. Rev. Lett.* **88**, 148301 (2002).
6. Tobar, M. E. & Blair, D. G. Sensitivity analysis of a resonant-mass gravitational wave antenna with a parametric transducer. *Rev. Sci. Instrum.* **66**, 2751–2759 (1995).
7. Long, J. C. *et al.* Upper limits to submillimetre-range forces from extra space-time dimensions. *Nature* **421**, 922–925 (2003).
8. White, J. D. An ultra high resolution displacement transducer using the Coulomb blockade electrometer. *Jap. J. Appl. Phys.* **32**, L1571–L1573 (1993).
9. Blencowe, M. P. & Wybourne, M. N. Sensitivity of a micromechanical displacement detector based on the radio-frequency single-electron transistor. *Appl. Phys. Lett.* **77**, 3845–3847 (2000).
10. Zhang, Y. & Blencowe, M. P. Intrinsic noise of a micro-mechanical displacement detector based on the radio-frequency single-electron transistor. *J. Appl. Phys.* **91**, 4249–4255 (2002).
11. Knobel, R. & Cleland, A. N. Piezoelectric displacement sensing with a single-electron transistor. *Appl. Phys. Lett.* **81**, 2258–2260 (2002).
12. Caves, C. M., Thorne, K. S., Drever, R. W. P., Sandberg, V. D. & Zimmermann, M. On the measurement of a weak classical force coupled to a quantum-mechanical oscillator. 1. Issues of principle. *Rev. Mod. Phys.* **52**, 341–392 (1980).
13. Huang, X. M. H., Zorman, C. A., Mehregany, M. & Roukes, M. L. Nanoelectromechanical systems: Nanodevice motion at microwave frequencies. *Nature* **421**, 496 (2003).
14. Abramovici, A. *et al.* Improved sensitivity in a gravitational wave interferometer and implications for LIGO. *Phys. Lett. A* **218**, 157–163 (1996).
15. Mamin, H. & Rugar, D. Sub-attoneutron force detection at millikelvin temperature. *Appl. Phys. Lett.* **79**, 3358–3360 (2001).
16. Cleland, A. N. & Roukes, M. L. External control of dissipation in a nanometer-scale radiofrequency mechanical resonator. *Sensors Actuators A* **72**, 256–261 (1999).
17. Beck, R. G. *et al.* GaAs/AlGaAs self-sensing cantilevers for low temperature scanning probe microscopy. *Appl. Phys. Lett.* **73**, 1149–1151 (1998).
18. Schoelkopf, R. J., Wahlgren, P., Kozhevnikov, A. A., Delsing, P. & Prober, D. E. The radiofrequency single-electron transistor (rf-SET): A fast and ultrasensitive electrometer. *Science* **280**, 1238–1242 (1998).
19. Devoret, M. H. & Schoelkopf, R. J. Amplifying quantum signals with the single-electron transistor. *Nature* **406**, 1039–1046 (2000).
20. Cleland, A. N., Aldridge, J. S., Driscoll, D. C. & Gossard, A. C. Nanomechanical displacement sensing using a quantum point contact. *Appl. Phys. Lett.* **81**, 1699–1701 (2002).
21. Fulton, T. A. & Dolan, G. J. Observations of single-electron charging effects in small tunnel junctions. *Phys. Rev. Lett.* **59**, 109–112 (1987).
22. Knobel, R., Yung, C. S. & Cleland, A. N. Single-electron transistor as a radio-frequency mixer. *Appl. Phys. Lett.* **81**, 532–534 (2002).
23. Martinis, J. M., Devoret, M. H. & Clarke, J. Experimental tests for the quantum behavior of a macroscopic degree of freedom: The phase difference across a Josephson junction. *Phys. Rev. B* **35**, 4682 (1987).
24. Greywall, D. S., Yurke, B., Busch, P. A., Pargellis, A. N. & Willett, R. A. Evading amplifier noise in nonlinear oscillators. *Phys. Rev. Lett.* **72**, 2992–2995 (1994).
25. Cleland, A. N. & Roukes, M. L. Fabrication of high frequency nanometer scale resonators from bulk Si crystals. *Appl. Phys. Lett.* **69**, 2653 (1996).
26. Yang, J., Ono, T. & Esashi, M. Surface effects and high quality factors in ultrathin single-crystal silicon cantilevers. *Appl. Phys. Lett.* **77**, 3860–3862 (2000).
27. Yasumura, K. Y. *et al.* Quality factors in micron- and submicron-thick cantilevers. *J. Microelectromech. Syst.* **9**, 117–125 (2000).

Acknowledgements We thank C. Yung, D. Schmidt and S. Aldridge for conversations, and B. Hill for processing support. We acknowledge support provided by the National Science Foundation XYZ-On-A-Chip programme, by the Army Research Office, and by the Office of Naval Research/DARPA SPINS programme.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.N.C. (cleland@physics.ucsb.edu).

Creep-strengthening of steel at high temperatures using nano-sized carbonitride dispersions

Masaki Taneike*, Fujio Abe* & Kota Sawada†

* Steel Research Center, and † Materials Information Technology Station, National Institute for Materials Science, 1-2-1 Sengen, Tsukuba 305-0047, Japan

Creep is a time-dependent mechanism of plastic deformation, which takes place in a range of materials under low stress—that is, under stresses lower than the yield stress¹. Metals and alloys can be designed to withstand creep at high temperatures, usually by a process called dispersion strengthening², in which fine particles are evenly distributed throughout the matrix. For example, high-temperature creep-resistant ferritic steels achieve optimal creep strength (at 923 K) through the dispersion of yttrium oxide nanoparticles³. However, the oxide particles are introduced by complicated mechanical alloying techniques and, as a result, the production of large-scale industrial components is economically unfeasible. Here we report the production of a 9 per cent Cr martensitic steel dispersed with nanometre-scale carbonitride particles using conventional processing techniques. At 923 K, our dispersion-strengthened material exhibits a time-to-rupture that is increased by two orders of magnitude relative to the current strongest creep-resistant steels⁴. This improvement in creep resistance is attributed to a mechanism of boundary pinning by the thermally stable carbonitride precipitates. The material also demonstrates enough fracture toughness. Our results should lead to improved grades of creep-resistant steels and to the economical manufacture of large-scale steel components for high-temperature applications.

Figure 1 shows the creep rate, corresponding to strain per hour, versus time curves of the 9% Cr steels with different carbon concentrations at 923 K and 140 MPa. The final data in the curves correspond to the ruptures of the respective specimens. The creep rate curves consist of the primary creep region, where the creep rate

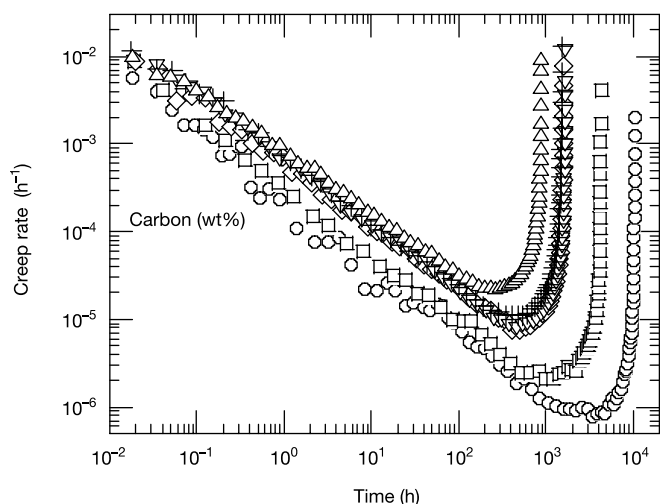


Figure 1 Creep rate versus time curves of at 923 K and 140 MPa. Circles, 0.002% C; squares, 0.018% C; diamonds, 0.047% C; inverse triangles, 0.078% C; crosses, 0.120% C; triangles, 0.160% C. The base composition of the steels was Fe–9% Cr–3% W–3% Co–0.2% V–0.06% Nb–0.05% N.

decreases with time, and the tertiary creep region, where the creep rate increases with time after reaching a minimum value. In the initial creep stage below 0.1 h, the creep rate is roughly the same for all specimens. The further decrease in creep rate with time and the onset time of tertiary creep are also the same among the specimens with higher carbon concentrations of 0.047 to 0.160%. However, the onset of tertiary creep is retarded up to longer times with decreasing carbon concentrations below 0.018%. The retardation of the onset of tertiary creep causes the longer duration of primary creep, which results in lower minimum creep rate and longer time to rupture. It is well known that the time to rupture is inversely proportional to the minimum creep rate¹. As shown in Fig. 2, the time to rupture is independent of carbon concentration in the higher carbon region above 0.047% but it significantly increases in the lower carbon region below 0.018%, which is also typical in applied tensile stresses within the range of 140–200 MPa.

Figure 3 compares the applied stress and time to rupture curves at 923 K among the present steel with 0.002% C and two conventional 9% Cr steels, 9% Cr–0.5% Mo–1.8% W–0.2% V–0.05% Nb and 9% Cr–1% Mo–0.2% V–0.05% Nb, which are specified as ASME Boiler and Pressure Vessel Code as grade P92 (ref. 5) and T91 (ref. 6), respectively. These two conventional steels contain about 0.08% C and 0.05% N. The P92 steel is known to possess the maximum creep strength of the ferritic heat-resistant steels for boiler components. The T91 steel has been widely used in power plants. The time to rupture of the 0.002% C steel presented here is approximately two orders of magnitude higher than the P92 steel and approximately three orders of magnitude higher than the T91 steel at 923 K and 140 MPa. The creep strength of the present steel is lower than that of ODS ferritic steel³. Furthermore, owing to its fine substructures of martensite, the 0.002% C steel exhibits ‘enough toughness’ values of 100–150 J at room temperature by the Charpy impact test. Enough toughness higher than 50 J is a basic requirement for thick-section structural components⁴. ODS ferritic steels are known to be brittle (as shown by 20 J or less) because of the ferrite matrix⁷.

A large number of fine precipitate particles having a size of 5–10 nm are distributed along prior austenite grain boundaries as well as along lath, block and packet boundaries in the specimen with 0.002% C after tempering heat treatment, as shown in Fig. 4. This is quite different from that in the specimen with 0.078% C, where mainly large particles of 100 to 300 nm are distributed together with the fine particles. The large particles were identified as M₂₃C₆ carbides rich in Cr (Supplementary Fig. S1). The fine particles

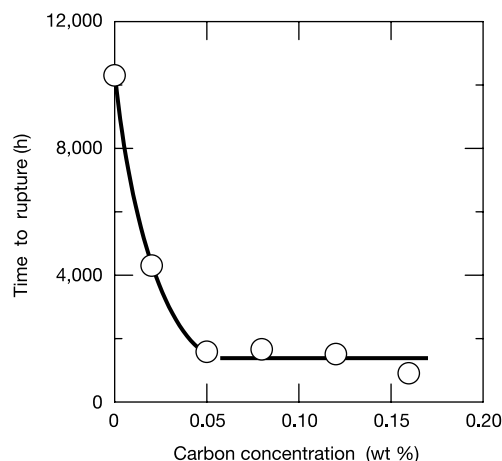


Figure 2 Time to rupture of the 9Cr steels at 923 K and 140 MPa, as a function of carbon concentration. The time to rupture corresponds to the time at the final data point in each curve in Fig. 1.

were identified as MX carbonitrides, where M means metallic alloying element and X is carbon and nitrogen (Supplementary Fig. S2). The MX carbonitrides were confirmed via energy dispersive microanalysis to be rich in V and Nb and also by energy filtered transmission electron microscopy based on electron energy loss spectroscopy (Supplementary Figs S2, S3 and S4). Any Cr-nitrides were not detected. The fine MX carbonitride particles are also distributed in the matrix within lath in addition to along boundaries, while the $M_{23}C_6$ carbide particles are distributed only along boundaries. Because the present steel contains Nb and V, which are stronger carbonitride-forming elements than Cr, at first Nb and V form carbonitrides upon cooling and during subsequent tempering. This consumes carbon that would be available for the formation of $M_{23}C_6$ carbides. As a result of the high volume fraction of MX precipitates, the volume fraction of $M_{23}C_6$ is reduced at low carbon concentrations. This causes a significant decrease in inter-particle distance and hence a significant increase in pinning force for migrating boundaries.

The coarsening rate of the MX carbonitride particles during creep at increased temperatures is much smaller than that of the $M_{23}C_6$ carbide particles⁸. This is mainly owing to the much lower solubility of V and Nb in the matrix than of Cr. The mean size of the MX carbonitride and $M_{23}C_6$ carbide particles in the present 9% Cr steels was measured to be 20–30 and 400 nm, respectively, along grain boundaries after creep for about 10^4 h at 923 K. It is known that the coarsening rate of V- and Nb-carbides is also very low—as low as the nitrides⁹. However, the addition of carbon to a 9% Cr steel causes the formation of a large amount of $M_{23}C_6$ carbides rich in Cr. Therefore, it is crucial for 9% Cr steels to reduce carbon content to very low amounts, so as to promote the formation of MX carbonitrides as very fine and thermally stable particles for prolonged periods of exposure at increased temperatures.

We have shown for martensitic 9% Cr steels that the migration of lath or block boundaries, causing the coarsening of lath or block, is closely correlated with the onset of tertiary creep and that the coarsening of lath or block by the migration of boundaries with absorbing excess dislocations is the major process in tertiary creep^{10,11}. The large number of nanometre-size MX carbonitride particles linking together along boundaries after heat treatment and their low coarsening rate during creep in the specimen with the very low carbon content of 0.002% suggest that large pinning forces for boundary migration is maintained for long times and that the

onset of tertiary creep is retarded up to long times. This effectively decreases the minimum creep rate and increases the time to rupture.

The Laves phase, Fe_2W , was also observed to have precipitated during creep at 923 K. The amount of Laves phase precipitated was approximately the same among the present steels with 0.002 to 0.16% C. Because the size of Laves phase particles was quite large—500–700 nm—and the number density was low after creep for 3.6×10^3 h at 923 K, precipitation strengthening due to the Laves phase is thought to be negligibly small. No evidence was found for the precipitation of the detrimental Z-phase, $Cr(Nb,V)N$, within the present test conditions. In a conventional 12% Cr steel with 0.29% Nb, 0.28% V and 0.074% N, massive particles of Z-phase form at the expense of fine MX carbonitrides and hence the precipitation of Z-phase causes a decrease of precipitation strengthening due to the MX carbonitrides¹². The lower concentrations of Z-phase-forming elements (such as 9% Cr, 0.05% Nb and 0.05% N) in the present 0.002% C steel than in the 12Cr steel suggest less pronounced formation of Z-phase.

A dispersion of fine carbonitrides was achieved in martensitic 9% and 12% Cr steels, but no creep strength data were given¹³. Their steels contained high nitrogen concentrations of 0.16 and 0.18% as well as high V concentrations of 0.66 and 0.76 wt% (ref. 13). Consequently, we prepared 9% Cr steels with high nitrogen concentrations of 0.074 and 0.103 wt% and with low carbon content of 0.002 wt%, and creep testing was carried out at 923 K. The results showed that the creep strength was rather lower for the high-nitrogen steels with 0.074 and 0.103% N than the present steel with 0.05% N. The creep strength is lower because the number density of undissolved large carbonitride particles increases with increasing nitrogen concentration. The MX carbonitride-forming

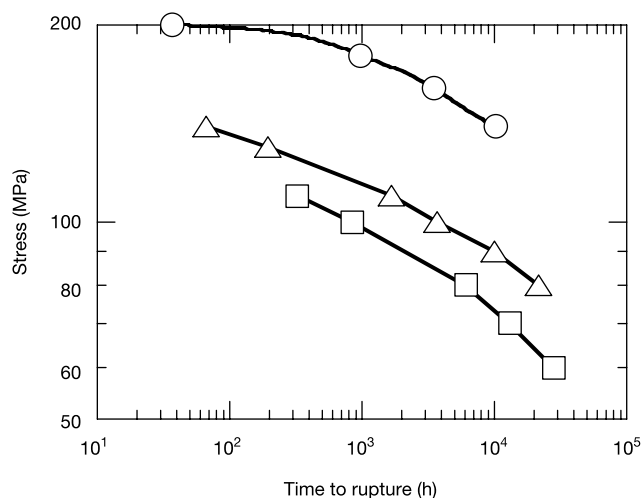


Figure 3 Stress and time to rupture relationship for the three 9% Cr steels at 923 K. Circles, the present steel with 0.002% C; triangles, 9% Cr–0.5% Mo–1.8% W–V–Nb steel specified as ASME-P92 (from ref. 5); squares, 9% Cr–1% Mo–V–Nb steel specified as ASME-T91 (from ref. 6).

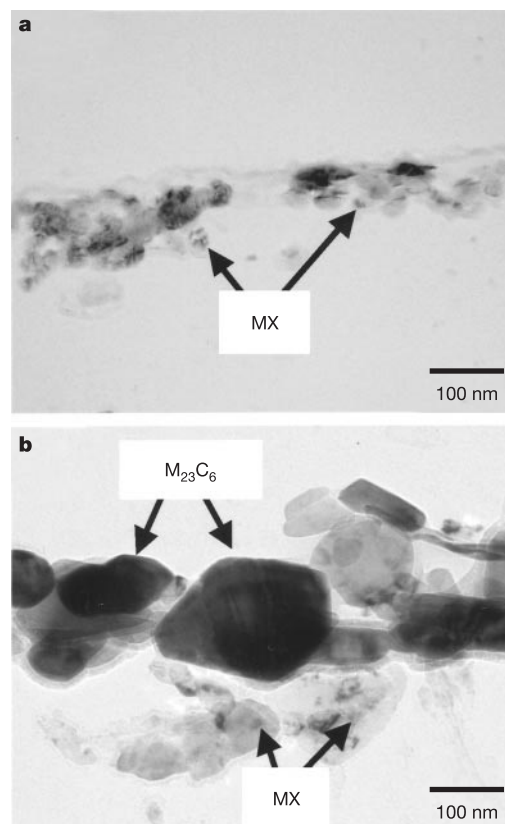


Figure 4 MX and $M_{23}C_6$ along grain boundaries in the specimens after heat treatment. **a**, 0.002% C steel and **b**, 0.078% C steel. Both steels were normalized at 1,373 K for 0.5 h, cooled in air and then tempered at 1,073 K for 1 h.

elements should be completely dissolved by appropriate heat treatments to maximize dispersion strengthening and so that complete dissolution is obtained at the relatively low nitrogen concentration of 0.05%. The addition of 3% W is also effective for the reduction of coarsening rate of MX carbonitrides during creep at 923 K (Supplementary Fig. S5).

The creep rupture strength for 10^5 h at operating temperature is a deciding criterion for the creep resistance of boiler and turbine components in power plants. The creep rupture strength at 923 K and 10^5 h was estimated using the time–temperature–parameter method with the Larson–Miller parameter¹⁴ at 100 MPa for the 0.002% C steel and 71 MPa for ASME-P92 (Supplementary Fig. S6).

We conclude that a dispersion of nanometre-size MX carbonitride particles is achieved for the martensitic 9% Cr steel by reducing carbon concentration to 0.002% and that their low coarsening rate at increased temperatures gives rise to excellent creep strength at 923 K, as shown by a time to rupture two orders of magnitude longer than for ASME-P92, which is the strongest of conventional steels for boiler components. □

Methods

The compositions of the steels examined were based on Fe–9% Cr–3% W–3% Co–0.2% V–0.06% Nb–0.05% N with 0.002, 0.018, 0.047, 0.078, 0.120 and 0.160 wt% C. The steels were prepared by vacuum induction melting to 50-kg ingots. Hot forging and hot rolling were performed to produce rods of 20 mm in diameter. The rods were normalized at 1,373 K for 0.5 h, cooled in air and then tempered at 1,073 K for 1 h. High-temperature strength was examined by creep tests at 923 K for up to about 10^4 h under constant load condition, using specimens of 10 mm in gauge diameter and 50 mm in gauge length. The specimens were manufactured according to ref. 15, which is substantially identical to ref. 16. During creep testing, an increase in strain with time was measured under constant tensile load. The applied load corresponded to the initial tensile stress of 200, 180, 160 and 140 MPa. The distribution of precipitates was observed by transmission electron microscope. The Charpy impact test was carried out at room temperature, using specimens of $10 \times 10 \times 55$ mm containing a 2-mm 45° V-notch with a root radius of 0.25 mm. The specimens were manufactured according to ref. 17.

Received 25 July 2002; accepted 14 May 2003; doi:10.1038/nature01740.

- Evans, R. W. & Wilshire, B. *Creep of Metals and Alloys* 1–308 (The Institute of Metals, London, 1985).
- Strudel, J.-L. in *Physical Metallurgy* (eds Cahn, R. & Haasen, P.) Ch. 22, 1412–1486 (North-Holland, Amsterdam/Oxford/New York/Tokyo, 1983).
- Ukai, S., Nishida, T., Okuda, T. & Yoshitake, T. R&D of oxide dispersion strengthened ferritic martensitic steels for FBR. *J. Nucl. Mater.* **258/263**, 1745–1749 (1998).
- Bakker, W. T. & Nath, B. in *Proc. 3rd EPRI Conference on Advances in Materials Technology for Fossil Power Plants*, Swansea, UK 1–4 (The Institute of Materials, London, 2001).
- National Institute for Material Science *Creep Data Sheet*. No. 48 (National Research Institute for Metals, Tokyo, 2002).
- National Research Institute for Metals *Creep Data Sheet*. No. 43 (National Research Institute for Metals, Tokyo, 1996).
- Ukai, S. *et al.* Tube manufacturing and mechanical properties of oxide dispersion strengthened ferritic steel. *J. Nucl. Mater.* **204**, 74–80 (1993).
- Sawada, K., Kubo, K. & Abe, F. Creep behaviour and stability of MX precipitates at high temperature in 9Cr–0.5Mo–1.8W–VNb steel. *Mater. Sci. Eng. A* **319/321**, 784–787 (2001).
- Wey, M. Y., Sakuma, T. & Nishizawa, T. Growth of alloy carbide particles in austenite. *Trans. Jpn. Inst. Met.* **22**, 733–742 (1981).
- Abe, F., Nakazawa, S., Araki, H. & Noda, T. The role of microstructural instability on creep behavior of a low radioactivation martensitic 9Cr–2W steel. *Metall. Trans. A* **23**, 469–477 (1992).
- Abe, F. Effect of quenching, tempering and cold rolling on creep deformation behavior of a tempered martensitic 9Cr–1W steel. *Metall. Mater. Trans. A* **34**, 913–925 (2003).
- Strang, A. & Vodarek, V. Z phase formation in martensitic 12CrMoVNb steel. *Mater. Sci. Technol.* **12**, 552–556 (1996).
- Goecmen, A., Steins, R., Solenthaler, C., Uggowitzer, P. J. & Speidel, M. O. Precipitation behaviour and stability of nitrides in high nitrogen martensitic 9% and 12% chromium steels. *ISIJ Int.* **36**, 768–776 (1996).
- Goldhoff, R. T. Some Observations on the extrapolation of high-temperature ferritic steel data. *Trans. ASME* **82D**, 848–854 (1960).
- Method of Creep and Creep Rupture Test for Metallic Materials*. Japanese Industrial Standard JIS Z 2271 (1999).
- Metallic Materials—Uninterrupted Uniaxial Creep Testing in Tension—Method of Test*. International Standard ISO 204 (Japanese Standards Association, Tokyo, 1997).
- Standard Methods for Notched Bar Impact Testing of Metallic Materials*. ASTM–E23 (American Society for Testing and Materials, Philadelphia, 1977).

Supplementary Information accompanies the paper on www.nature.com/nature.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.A. (ABE.Fujio@nims.go.jp).

Measuring dipolar width across liquid–liquid interfaces with ‘molecular rulers’

William H. Steel & Robert A. Walker

Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742, USA

Molecular dynamics simulations have previously described how the physical properties across immiscible liquid–liquid interfaces should converge from aqueous to organic limits^{1–5}, but these predictions have largely gone untested, owing to difficulties associated with probing buried interfaces. X-ray and neutron scattering experiments have created detailed pictures of molecular structure at these boundaries^{6–8}, but such scattering studies cannot probe how surface-altered solvent structures affect interfacial solvating properties. Given that surface-mediated solvent properties control interfacial solute concentrations and reactivities, identifying the characteristic dimensions of interfacial solvation is essential for formulating predictive models of solution phase surface chemistry. Here we use specially synthesized solvatochromic surfactants that act as ‘molecular rulers’⁹ and resonance-enhanced second-harmonic generation^{10–13} to measure the dipolar width of weakly and strongly associating liquid–liquid interfaces. Dipolar width describes the distance required for a dielectric environment to change from one phase to another. Our results show that polarity converges to a non-polar limit on subnanometre length scales across a water–cyclohexane interface. However, polarity across the strongly associating, water–1-octanol interface is dominated by a non-polar, alkane-like region. These data call into question the use of continuum descriptions of liquids to characterize interfacial solvation, and demonstrate that interfacial environments can vary in a non-additive manner from bulk solution limits.

Measuring dipolar width requires systematic changing of the distribution of a probe relative to a nominal interfacial dividing surface. We accomplish this task using newly created surfactants known as ‘molecular rulers’⁹. Molecular ruler surfactants consist of anionic sulphate groups attached to hydrophobic, solvatochromic probes by different length alkyl spacers (Fig. 1). The probes themselves are based on *p*-nitroanisole, an aromatic solute whose bulk solution excitation wavelength (λ_{max}) monotonically shifts by more than 20 nm as solvent polarity, as characterized by the static dielectric constant ϵ , changes from that of cyclohexane ($\epsilon = 2.0$, $\lambda_{\text{max}} = 293$ nm) to that of water ($\epsilon = 78.9$, $\lambda_{\text{max}} = 316$ nm)¹⁴. Partitioning experiments show *p*-nitroanisole to be approximately 20 times more soluble in cyclohexane than water. Assuming that the anionic headgroup remains solvated in the aqueous phase, longer alkyl spacers should allow the aromatic probe to penetrate further into the organic phase. Monitoring probe excitation wavelength as a function of alkyl spacer length measures the distance required for solvent polarity to change from an aqueous to an organic limit.

To acquire excitation spectra of these probes at different liquid–liquid interfaces, we use resonance-enhanced second-harmonic generation (SHG), a second-order, nonlinear optical method that is both surface- and molecule-specific^{10–13}. Within the electric dipole approximation, the SHG response is restricted to anisotropic environments. Liquids themselves are isotropic, meaning that only those solutes at a liquid–liquid interface are responsible for the observed SHG signal. In a typical SHG experiment, a single coherent optical field with frequency ω is focused on the interface under study, and a nonlinear polarization with frequency 2ω is detected. The intensity of the 2ω field is proportional to the square

of the second-order susceptibility, $\chi^{(2)}$:

$$I(2\omega) \propto |\chi^{(2)}|^2 I^2(\omega) \quad (1)$$

where $I(\omega)$ is the intensity of the incident field, and $\chi^{(2)}$ is a third-rank tensor that is antisymmetric with respect to inversion. The $\chi^{(2)}$ tensor is responsible for the technique's inherent surface specificity, and contains both nonresonant and resonant contributions:

$$\chi^{(2)} = \chi_{\text{nr}}^{(2)} + \chi_{\text{r}}^{(2)} \quad (2)$$

Typically, the resonant term is much larger than the nonresonant contribution and can be related to microscopic hyperpolarizability¹⁵:

$$\chi_{\text{r}}^{(2)} = \sum_{k,e} \frac{\langle \mu_{\text{gk}} \mu_{\text{ke}} \mu_{\text{eg}} \rangle}{(\omega_{\text{gk}} - \omega - i\Gamma)(\omega_{\text{eg}} - 2\omega + i\Gamma)} \quad (3)$$

where μ_{ij} is the transition matrix element between states i and j (here g stands for ground state, k for an intermediate, virtual state, and e for the first excited state) and Γ is a linewidth term. When 2ω is resonant with ω_{eg} , $\chi^{(2)}$ becomes large, leading to a strong resonance enhancement in the observed intensity at 2ω . Thus, measuring the scaled intensity $[I(2\omega)/I^2(\omega)]$ as a function of 2ω records an effective excitation spectrum of solutes adsorbed to an interface^{16,17}.

Figure 2 compares SHG spectra of a neutral chromophore—*p*-nitroanisole—and three molecular rulers— C_2 , C_4 and C_6 —adsorbed to a weakly associating, water–cyclohexane interface. Surface excess concentrations of all surfactants (including *p*-nitroanisole) varied between 0.5×10^{13} and 1.5×10^{13} molecules cm^{-2} , as determined by surface pressure isotherms acquired using the Wilhelmy plate method⁹. With each system, bulk aqueous surfactant concentrations were adjusted to keep surface coverages at $<10\%$ of a full monolayer and minimize any possible chromophore–chromophore interactions. Such low surface concentrations lead to only slight variation in interfacial tension ($\sim 2\text{--}5 \text{ mN m}^{-1}$), supporting the claim that the molecular rulers probe interfacial properties, rather than properties of an interface in the presence of saturated, charged monolayers. It is important to note that these experiments probe the environment surrounding a solute at the interface, not the neat interface itself. The dipolar properties across a boundary between two pure solvents may change in the absence of dilute, adsorbed surfactants.

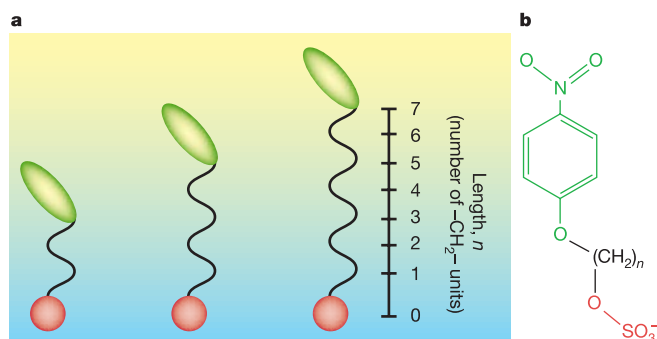


Figure 1 Schematic representation of adsorbed molecular ruler surfactants, and their general structure. **a**, Schematic representation of molecular ruler surfactants adsorbed to a liquid–liquid interface. As the alkyl spacer between the headgroup (red circle) and solvatochromic chromophore (green ellipse) lengthens, the hydrophobic probe can extend further into the organic phase. Correlating ruler length (in $-\text{CH}_2-$ units) with chromophore excitation wavelength enables experiments to determine the distance required for interfacial solvent polarity to converge to the organic limit. **b**, General structure of molecular ruler surfactants. Surfactants are referred to as C_n rulers, where n corresponds to the number of methylene ($-\text{CH}_2-$) groups in the alkyl spacer. Excitation wavelengths of the hydrophobic probe in different solvents are noted in the text.

In aqueous solution, chromophores exhibit excitation maxima near $318 \pm 2 \text{ nm}$, while in bulk cyclohexane, excitation maxima are centred around $295 \pm 2 \text{ nm}$. The SHG spectrum of *p*-nitroanisole adsorbed to the water–cyclohexane interface exhibits a maximum at 308 nm . This result is consistent with models of interfacial polarity suggesting that the local dielectric environment between two immiscible liquids can be represented by averaged contributions from the adjacent phases^{16–18}. Simulations suggest that local polarity around an adsorbed solute should be very sensitive to solute location as well as solute orientation relative to a nominal interfacial dividing surface¹⁹. Results reported elsewhere²⁰ show how subtle changes in solute functionality can lead to marked variations in the interfacial polarity sampled by different solutes at the same interface. However, experimental results do not identify the distance required for polarity to converge to a bulk, organic limit.

The C_2 ruler separates the hydrophobic, *p*-nitroanisole chromophore from the hydrophilic, anionic sulphate by two methylene groups or $\sim 3 \text{ \AA}$. SHG data from the C_2 ruler adsorbed to the water–cyclohexane interface shifts seven nanometres to shorter wavelengths relative to the *p*-nitroanisole SHG spectrum. The spectrum also broadens considerably. The wavelength shift in chromophore excitation indicates a less polar environment relative to the one sampled by *p*-nitroanisole at the same water–cyclohexane interface. Spectral broadening can result from a distribution of chromophore orientations and/or solvation environments. Polarization-dependent SHG measurements do not support orientational effects as the origin of spectral broadening. Using methods described elsewhere^{11,13}, we determined the orientations of the C_2 , C_4 and C_6 ruler chromophores to be 47° , 51° and 45° relative to the surface normal, respectively. To within the experimental uncertainty ($\pm 5^\circ$), these results are indistinguishable. Were the longer ruler to adopt an orientation parallel to the interface, we would expect larger deflec-

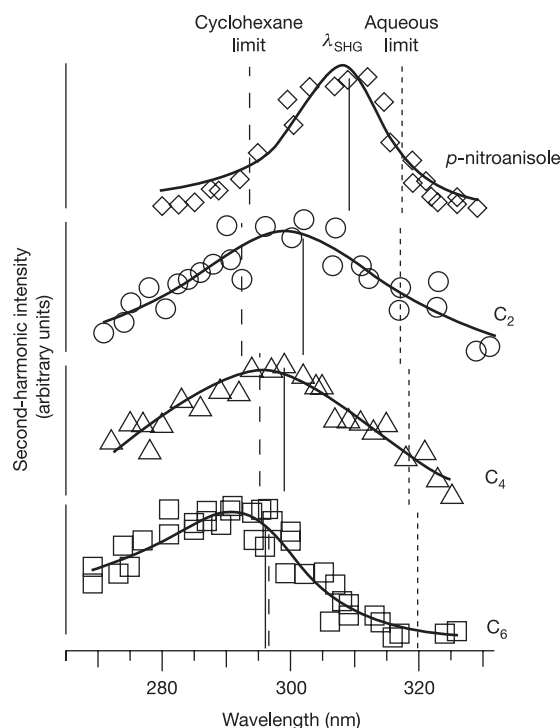


Figure 2 Resonance-enhanced SHG spectra of (top to bottom) *p*-nitroanisole, C_2 rulers, C_4 rulers and C_6 rulers adsorbed to a water–cyclohexane interface. Dashed and dotted lines denote excitation maxima in bulk cyclohexane and water, respectively. Solid vertical lines correspond to SHG maxima (λ_{SHG}) as determined by fitting the data to equations (1)–(3). Note that SHG maxima do not always correspond to the wavelengths with highest SHG intensity, owing to the nonresonant contribution to $\chi^{(2)}$ in equation (2).

tions from the surface normal. Large variations in local polarity are consistent with the effects of capillary wave action and with neutron scattering studies that show distributions in surfactant 'float depth' across an air–water interface^{21,22}. Consequently, we attribute the width of the C₂ ruler spectrum to the broad distribution in local polarity resulting from surface roughness.

The chromophore attached to the C₆ ruler samples an environment that is clearly nonpolar, as evidenced by a SHG spectrum centred at 296 nm. The C₆ ruler SHG spectrum also narrows relative to that of the C₂ ruler, suggesting a more uniform local polarity. A six-carbon alkyl chain stretches less than 9-Å—or three water diameters—when fully extended. This value sets an upper limit on the distance across which solvent polarity converges from an aqueous to an alkane limit. Should the chain contain conformational defects, or should the ruler adopt an orientation other than normal to the surface, interfacial width will be less than the 9-Å length of the fully extended ruler. Given the short alkyl chain lengths and supporting optical⁹ and X-ray scattering studies²³ as well as molecular dynamics simulations²⁴, we anticipate that there should be little conformational disorder in the molecular rulers, especially when adsorbed to a water–organic interface. The water–cyclohexane results suggest a molecularly sharp dipolar width, and support recent scattering studies and simulations of solvent density profiles across weakly associating liquid–liquid interfaces^{5,8}.

In contrast to the weakly associating water–cyclohexane interface, the water–1-octanol interface exhibits a low interfacial tension (8.5 mN m⁻¹) that reflects strong hydrogen bonding between the two protic but immiscible liquids. Figure 3 shows the SHG spectra acquired from C₂, C₄, C₆ and C₈ molecular rulers adsorbed to the water–1-octanol interface. Again, aqueous concentrations were adjusted to keep surface excess concentrations at <10% of full monolayer coverage, approximately $(0.3\text{--}0.5) \times 10^{13}$ molecules cm⁻² at the water–1-octanol interface.

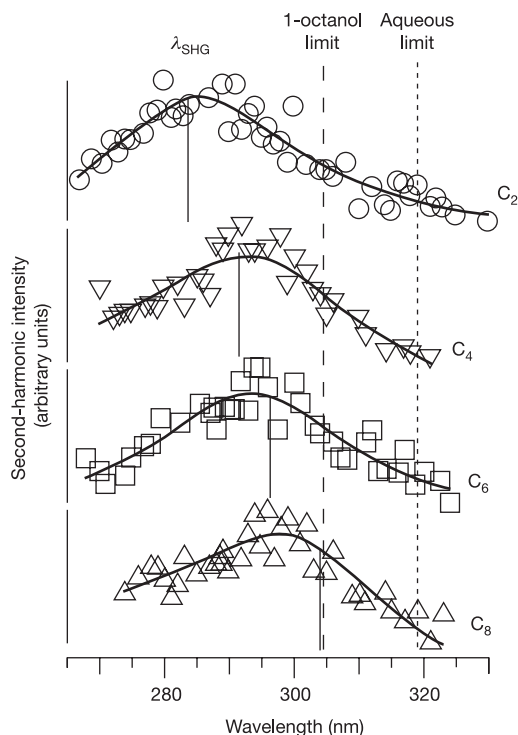


Figure 3 Resonance-enhanced SHG spectra of (top to bottom) C₂ rulers, C₄ rulers, C₆ rulers and C₈ rulers adsorbed to a water–1-octanol interface. Dashed and dotted lines denote excitation maxima in bulk octanol and water, respectively. Solid vertical lines correspond to SHG maxima as determined by fitting the data to equations (1)–(3).

Striking differences exist between SHG spectra of rulers adsorbed to weakly and strongly associating interfaces. With a static dielectric constant of 10.3 (20 °C), 1-octanol is considerably more polar than cyclohexane ($\epsilon = 2.0$, 20 °C). Consequently, the excitation window for the water–1-octanol system extends only 15 nm from 318 nm (in bulk water) to 303 nm (in bulk octanol). Nevertheless, the SHG spectrum of the C₂ ruler adsorbed to the water–1-octanol interface shows a strong maximum at 285 ± 2 nm. For comparison, λ_{max} of *p*-nitroanisole dissolved in *n*-octane is 292 nm. The SHG maximum shifts to longer wavelengths with increasing ruler length. Interfacial and bulk organic polarity converge in the spectrum of a C₈ ruler adsorbed to the water–1-octanol interface. Chromophore orientations of all four molecular rulers adsorbed to the water–1-octanol interface vary between 36° and 39° relative to the surface normal. Smaller, more uniform chromophore deflections at the water–1-octanol interface imply that strong, interfacial interactions lead to more pronounced polar ordering of adsorbed solutes.

On the basis of probe solvatochromic behaviour, we conclude that the C₂ molecular ruler adsorbed to the water–1-octanol interface samples an interfacial polarity that is considerably less than that of either bulk phase. As the molecular ruler lengthens, the local dielectric environment around the probe converges to bulk solution polarity from the nonpolar limit. Similar behaviour has been observed across strongly associating solid–liquid interfaces where long, *n*-alcohol species create environments that are less polar than bulk solution²⁵. This effect diminishes as alcohol chain length shortens, and had been attributed to surface-induced polar ordering of the interfacial solvent species. If strong hydrogen bonding between the aqueous and 1-octanol phases leads to monolayer-like structure within the first solvent layer, then the local interfacial polarity sampled by the ruler chromophores could very well approach an alkane limit.

Such a picture is consistent with the formation of moderately dense monolayers of long-chain alcohols recently observed at water–hexane interfaces²⁶. Molecular rulers would have to be longer than the organic solvent in order for the probe to extend beyond the surface-altered nonpolar region. Assuming that the anionic sulphate group remains solvated in the aqueous phase, geometric considerations place an upper limit on this interfacial distance of ~ 1.2 nm, the length of a fully extended, C₈ ruler. The length of an extended 1-octanol molecule is also ~ 1.2 nm. (Experiments described in this Letter are sensitive to the separation between the charged headgroup and the hydrophobic, solvatochromic probe. Experiments cannot distinguish absolute headgroup position relative to the nominal, interfacial boundary. However, simulations²⁴ predict that except for the ether oxygen, the sulphate group of sodium dodecyl sulphate adsorbed to a liquid–liquid interface is solvated by the aqueous subphase. Furthermore, scattering data²¹ show that a quaternized, trimethyl ammonium headgroup (of C₁₆H₃₃N(CH₃)₃⁺) is solvated in an aqueous phase up to the nitrogen. The sulphate anion in molecular rulers is considerably less hydrophobic than a quaternized ammonium, thus we anticipate that the molecular ruler headgroup will remain solvated primarily by water.)

An interesting aspect of the data in Fig. 3 is their implication for the equilibrium that exists across the water–1-octanol interface. At room temperature, saturated 1-octanol phase is 5% water by weight²⁷. However, spectra in Fig. 3 clearly show that some portion of the interfacial boundary is nonpolar, meaning that despite the flux of water across the interface in both directions, there exists a region of alkane-like polarity. Based on temperature-dependent solubility data, the enthalpy associated with solvating water in octanol is 10 kJ mol⁻¹ (ref. 27). This energetically unfavourable result is equivalent to the loss of approximately half of a hydrogen bond, but does not identify the magnitude of the barrier to water passage across the nonpolar region. Should the nonpolar region sampled by ruler chromophores mimic that of small *n*-alkanes,

water-in-octane solubility data²⁸ place a lower limit on this barrier of 40 kJ mol⁻¹.

Results from the water–1-octanol system also provide insight into methods used to characterize the membrane affinity of different pharmaceutical and natural products. The ‘logP’ scale describes a solute’s tendency to partition across a water–1-octanol interface, and has been used for almost 70 years as a measure of biological activity²⁹. This scale evolved from empirical searches for systems that correlated with known physiological efficacy of different anaesthetics. The spectra in Fig. 3 imply that the success of the ‘logP’ scale may, in part, be due to a membrane-like structure induced in the interfacial octanol solvent by the underlying aqueous phase.

The present experiments have used a family of surfactants to determine how solvent polarity changes across weakly and strongly associating liquid–liquid interfaces. Using resonance-enhanced SHG, we measure probe excitation wavelengths as a function of probe–headgroup separation. Solvent polarity across the weakly associating, water–cyclohexane interface smoothly converges from the aqueous to the organic limit in less than 9 Å. This result supports existing models of molecularly sharp interfaces proposed by simulations and X-ray scattering results. The strongly associating, water–1-octanol interface presents a region that is considerably less polar than either bulk phase, implying that surface-induced ordering of the organic solvent creates a hydrophobic barrier between the two polar liquids. □

Received 20 February; accepted 28 May 2003; doi:10.1038/nature01791.

1. Viceli, J. & Benjamin, I. Adsorption at the interface between water and self-assembled monolayers: Structure and electronic spectra. *J. Phys. Chem. B* **106**, 7898–7907 (2002).
2. Chang, T. M. & Dang, L. X. Molecular dynamics simulations of CCl₄–H₂O liquid–liquid interface with polarizable potential models. *J. Chem. Phys.* **104**, 6772–6783 (1996).
3. Chipot, C., Wilson, M. A. & Pohorille, A. Interactions of anesthetics with the water–hexane interface. A molecular dynamics study. *J. Phys. Chem. B* **101**, 782–791 (1997).
4. DaRocha, S. R. & Rossky, P. J. Surfactant modified CO₂–water interface: A molecular view. *J. Phys. Chem. B* **106**, 13250–13261 (2002).
5. Senapati, S. & Berkowitz, M. L. Computer simulation study of the interface width of the liquid/liquid interface. *Phys. Rev. Lett.* **87**, 176101 (2001).
6. Lee, L. T., Langevin, D. & Farnoux, B. Neutron reflectivity at liquid interfaces. *Physica B* **198**, 83–88 (1994).
7. Penfold, J., Richardson, R. M., Zarbakhsh, A. & Webster, J. R. P. Recent advances in the study of chemical surfaces and interfaces by specular neutron reflection. *J. Chem. Soc. Faraday Trans.* **93**, 3899–3917 (1997).
8. Tikhonov, A. M., Mitrinovic, D. M., Li, M., Huang, Z. & Schlossman, M. L. An X-ray reflectivity study of the water–docosane interface. *J. Phys. Chem. B* **104**, 6336–6339 (2000).
9. Steel, W. H., Damkaci, E., Nolan, R. & Walker, R. A. Molecular rulers: New families of molecules for measuring interfacial widths. *J. Am. Chem. Soc.* **124**, 4824–4831 (2002).
10. Antoine, A., Bianchi, F., Brevet, P. F. & Girault, H. H. Studies of water/alcohol and air/alcohol interfaces by second harmonic generation. *J. Chem. Soc. Faraday Trans.* **93**, 3833–3838 (1997).
11. Corn, R. M. & Higgins, D. A. Optical second harmonic spectroscopy as a probe of surface chemistry. *Chem. Rev.* **94**, 107–125 (1994).
12. Eienthal, K. B. Photochemistry and photophysics of liquid interfaces by second harmonic spectroscopy. *J. Phys. Chem.* **100**, 12997–13006 (1996).
13. Zhuang, X., Miranda, P. B., Kim, D. & Shen, Y. R. Mapping molecular orientation and conformation at interfaces by surface nonlinear optics. *Phys. Rev. B* **59**, 12632–12640 (1999).
14. Laurence, C., Nicolet, P. & Dalati, M. T. The empirical treatment of solvent–solute interactions: 15 years of π^* . *J. Phys. Chem.* **98**, 5807–5816 (1994).
15. Shen, Y. R. Surface properties probed by second harmonic and sum frequency generation. *Nature* **337**, 519–525 (1989).
16. Wang, H., Borguet, E. & Eienthal, K. B. Polarity of liquid interfaces by second harmonic generation spectroscopy. *J. Phys. Chem. A* **101**, 713–718 (1997).
17. Wang, H., Borguet, E. & Eienthal, K. B. Generalized interface polarity scale based on second harmonic spectroscopy. *J. Phys. Chem. B* **102**, 4927–4932 (1998).
18. Ishizaka, S., Kim, H. B. & Kitamura, N. Time-resolved total internal reflection fluorometry study on polarity at a liquid/liquid interface. *Anal. Chem.* **73**, 2421–2428 (2001).
19. Benjamin, I. Solvent effects on electronic spectra at liquid interfaces. A continuum electrostatic model. *J. Phys. Chem. A* **102**, 9500–9506 (1998).
20. Steel, W. H. & Walker, R. A. Solvent polarity at an aqueous/alkane interface: The effect of solute identity. *J. Am. Chem. Soc.* **125**, 1132–1133 (2003).
21. Li, Z. X., Bain, C. D., Thomas, R. K., Duffy, D. C. & Penfold, J. Monolayers of hexadecyltrimethylammonium p-tosylate at the air–water interface 2. Neutron reflection. *J. Phys. Chem. B* **102**, 9473–9480 (1998).
22. Penfold, J. & Thomas, R. K. Solvent distribution in non-ionic surfactant monolayers. *Phys. Chem. Chem. Phys.* **4**, 2648–2652 (2002).
23. Tikhonov, A. M. & Schlossman, M. L. Surfactant and water ordering in triacontanol monolayers at the water–hexane interface. *J. Phys. Chem. B* **107**, 3344–3347 (2003).
24. Schweighofer, K., Essmann, U. & Berkowitz, M. Simulation of sodium dodecyl sulfate at the water–

vapor and water–carbon tetrachloride interfaces at low surface coverages. *J. Phys. Chem. B* **101**, 3793–3799 (1997).

25. Zhang, X. & Walker, R. A. Discrete partitioning of solvent permittivity at liquid–solid interfaces. *Langmuir* **17**, 4486–4489 (2001).
26. Zhang, Z., Mitrinovic, D. M., Williams, S. M., Huang, Z. & Schlossman, M. L. X-ray scattering from monolayers of F(CF₂)₁₀(CH₂)₂OH at the water–(hexane solution) and water–vapor interfaces. *J. Chem. Phys.* **110**, 7421–7432 (1999).
27. Barton, A. F. M. (ed.) *IUPAC Solubility Data Series: Alcohols with Water* (Pergamon, Oxford, 1984).
28. Tsionopoulos, C. & Wilson, G. M. High-temperature mutual solubilities of hydrocarbons and water. *Am. Inst. Chem. J.* **31**, 376–384 (1983).
29. Sangster, J. *Octanol–Water Partition Coefficients* (ed. Fogg, P. G. T.) 2–15 (Wiley and Sons, New York, 1997).

Acknowledgements This work was supported by the Research Corporation and the National Science Foundation through its CAREER Program.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.A.W. (rw158@umail.umd.edu).

Palaeoceanographic implications of genetic variation in living North Atlantic *Neogloboquadrina pachyderma*

D. Bauch*, K. Darling†, J. Simstich*, H.A. Bauch*‡, H. Erlenkeuser§ & D. Kroon||

* GEOMAR, Wischhofstrasse 1–3, 24148-Kiel, Germany

† School of GeoSciences and Institute of Cell, Animal & Population Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JG, UK

‡ Mainz Academy of Sciences, Humanity and Literature, Geschwister-Scholl-Strasse 2, 55131-Mainz, Germany

§ Leibniz Laboratory, Kiel University, Max-Eyth Strasse, 24118-Kiel, Germany

|| Vrije Universiteit Amsterdam, De Boelelaan 1085, N-1081 HV Amsterdam, The Netherlands

The shells of the planktonic foraminifer *Neogloboquadrina pachyderma* have become a classical tool for reconstructing glacial–interglacial climate conditions in the North Atlantic Ocean^{1–3}. Palaeoceanographers utilize its left- and right-coiling variants, which exhibit a distinctive reciprocal temperature and water mass related shift in faunal abundance both at present and in late Quaternary sediments^{1,2,4,5}. Recently discovered cryptic genetic diversity in planktonic foraminifers^{6–8} now poses significant questions for these studies. Here we report genetic evidence demonstrating that the apparent ‘single species’ shell-based records of right-coiling *N. pachyderma* used in palaeoceanographic reconstructions contain an alternation in species as environmental factors change. This is reflected in a species-dependent incremental shift in right-coiling *N. pachyderma* shell calcite $\delta^{18}\text{O}$ between the Last Glacial Maximum and full Holocene conditions. Guided by the percentage dextral coiling ratio, our findings enhance the use of $\delta^{18}\text{O}$ records of right-coiling *N. pachyderma* for future study. They also highlight the need to genetically investigate other important morphospecies to refine their accuracy and reliability as palaeoceanographic proxies.

Morphological distinction provides the main basis for foraminiferal counts and derived palaeoceanographic reconstructions^{2,9,10}, and yet the genus *Neogloboquadrina* is known to be a rather variable morphogroup^{5,11–14}. Although lack of DNA in fossil foraminiferal shells precludes genetic investigation, molecular analyses of the

living *N. pachyderma* planktonic assemblages now provide the opportunity to resolve much of this confusion. The left-coiling ('sin.') form of the morphospecies *N. pachyderma* is the dominant morphotype today in the cold, high-northern latitudes^{15,16} (Fig. 1a). Although right-coiling ('dex.') *N. pachyderma* is found in more temperate environments than its left-coiled counterpart (Fig. 1a), a small percentage of *N. pachyderma* (dex.) persists in sedimentary records, even in regions with extreme polar conditions such as the central Arctic Ocean¹⁷. It is unknown whether these *N. pachyderma* (dex.) represent sporadic hydrographic change, expatriation from their more-temperate water mass, or whether they are aberrant forms of the left-coiling variety. Recently, the $\delta^{18}\text{O}$ signature in *N. pachyderma* (dex.) has been proposed as a good recorder of glacial-interglacial sea surface temperatures^{18–20}. Our molecular work now questions such continuous 'single species' use of the *N. pachyderma* (dex.) $\delta^{18}\text{O}$ record for past climate reconstruction. We have investigated this species-versus-morphology issue by comparing the genotypic identity of *N. pachyderma* (sin. and dex.) collected from the Nordic seas surface waters against a background of *N. pachyderma* $\delta^{18}\text{O}$ values and coiling ratios from the equivalent regional surface sediments. We then demonstrate the palaeoceanographic application of this relationship down-core.

The general modern hydrography of the Nordic seas is reflected in the stable oxygen isotope values of both *N. pachyderma* (sin.) and *N. pachyderma* (dex.)²¹, since the $\delta^{18}\text{O}$ incorporated into test calcite is a function of temperature and salinity²². The $\delta^{18}\text{O}$ values of *N. pachyderma* (sin.) and *N. pachyderma* (dex.) are largely identical under polar conditions (Fig. 1b) in the western region, where relative abundances of *N. pachyderma* (dex.) are low (Fig. 1a). But in the east and south of the Nordic seas, where the percentage of *N. pachyderma* (dex.) increases rapidly under more-temperate conditions (Fig. 1a), a significant isotopic difference between left- and right-coiling *N. pachyderma* of up to about 1‰ in $\delta^{18}\text{O}$ is observed (Fig. 1b).

Molecular analysis of the small subunit (SSU) ribosomal RNA sequences has previously shown that the *N. pachyderma* morpho-

species consists of a complex of distinct genotypes⁷, and that the coiling direction in the Southern Hemisphere is not a phenotypic response to temperature, but reflects genetic divergence and the probable existence of two separate species⁷. In this study (Fig. 2), we have found that throughout the Nordic seas, all left-coiling *N. pachyderma* are genotypically [Type I (sin.)]. This is not the case for *N. pachyderma* (dex.). In the region of the Norwegian Sea and Denmark Strait south of 64° N (Fig. 2), all *N. pachyderma* (dex.) are genotypically [Type I (dex.)] and are genetically highly distinct from the left-coiling variant⁷. Under polar conditions north of the latitude 70° N where the coiling ratios are confined to less than 5% (Fig. 1a) and the $\delta^{18}\text{O}$ values of *N. pachyderma* (sin.) and *N. pachyderma* (dex.) are largely the same (Fig. 1b), genotyping reveals that the left- and right-coiling variants of *N. pachyderma* are genotypically identical. Both are [Type I (sin.)]. Clearly, *N. pachyderma* [Type I (dex.)] is thriving in the relatively warm North Atlantic waters, and is adapted to a totally different hydrographic regime than left- and right-coiling [Type I (sin.)], which is dominant in polar waters. Such genetic and environmental evidence indicates that these two genetic types should be considered different species. As might be expected, there is a mixing zone between the right- and left-coiling provinces (70° N to 64° N) where the dextral morphotypes can be either *N. pachyderma* [Type I (dex.)] or [Type I (sin.)].

The genetic evidence for *N. pachyderma* shows that the morphological distinction of coiling directions is not sufficient to distinguish between genotypes. Since fossil specimens do not contain DNA for genetic analysis, we need an indicator proxy to distinguish between the two species of right-coiling *N. pachyderma*. From the presented sediment data in combination with the genetic evidence it can be assumed that in regions where the percentage of *N. pachyderma* (dex.) does not exceed a 'threshold' value of ~5% in coiling direction, both *N. pachyderma* coiling variants will be identical, genotypically [Type I (sin.)], with a correspondingly identical isotopic composition. In regions where the percentage of *N. pachyderma* (dex.) exceeds this 'threshold' value, *N. pachyderma*

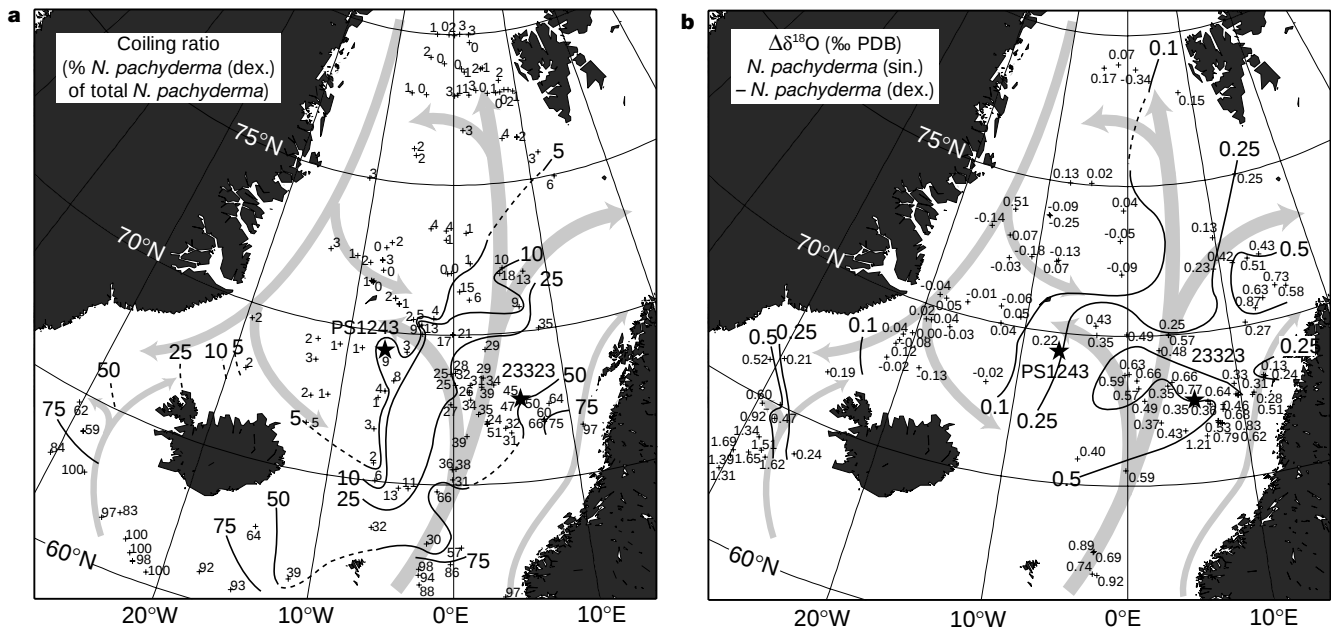


Figure 1 Spatial distributions of coiling ratios and isotopic differences. **a**, Coiling ratios of *N. pachyderma* in surface sediments of the Nordic seas based on faunal census data (>150 μm)¹⁰. The core locations of PS1243 and M23323 are highlighted. The arrows show the general surface ocean currents: the southward-flowing cold East Greenland Current with side branches; the northward-flowing warm Norwegian Current with side

branches; also indicated are the warm Norwegian Coastal and Irminger currents along the Norwegian and Iceland coasts, respectively. **b**, Isotopic difference between *N. pachyderma* (sin.) and *N. pachyderma* (dex.) from surface sediments in the Nordic seas²¹. For each analysis, 20–30 shells have been measured from the size fraction 125–250 μm .

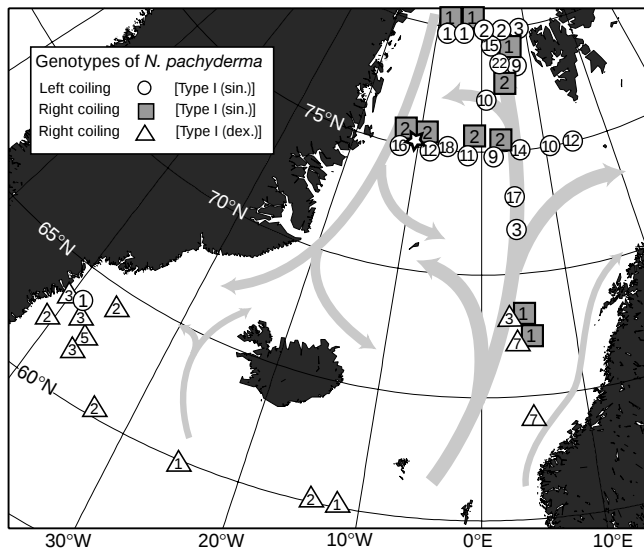


Figure 2 Geographical distribution of genotypes of left-coiling *N. pachyderma* [Type I (sin.)], the right-coiling form of *N. pachyderma* [Type I (sin.)], and right-coiling *N. pachyderma* [Type I (dex.)]. Data from the Denmark Strait region are taken from ref. 8. Numbers within symbols indicate number of specimens genotyped at the specific locations. At a single position, marked with a star, there was sufficient suitable material in a multinet to derive an isotope signature for both left- and right-coiling *N. pachyderma* from the same plankton sample. Identical $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ values were found for both, corroborating our evidence that in this regional water mass they are the same genetic type.

(dex.) is expected to have a progressively different isotopic composition from *N. pachyderma* (sin.) as the *N. pachyderma* [Type I (dex.)] genotype increasingly dominates the assemblage.

In order to further test our conclusions from the modern data set, we investigated two sediment cores representing the warmer and colder regions of the Norwegian Sea (see Fig. 1 for locations). Samples from the glacial sections of the two cores show a low relative abundance (1–5%) of *N. pachyderma* (dex.) (Fig. 3). Since the early Holocene (~9–10 kyr ago), dextral coiling ratios are higher, with levels of 5–10% in core PS1243 and 40–60% in core M23323 due to climate warming at the end of the deglaciation. These faunal differences reflect the basic water-mass temperature gradient between the two sites. A similar picture is recognized in the stable oxygen isotope records of *N. pachyderma* (sin.) and *N. pachyderma* (dex.), which both depict the overall climatic change since the Last Glacial Maximum (LGM) (Fig. 3). However, while $\delta^{18}\text{O}$ in both species are nearly the same during the glacial sections, when dextral coiling ratios are extremely low, a systematic offset of about 0.5‰ is noted since about 10 kyr ago when dextral coiling ratios increase, accompanied by a decrease in the relative abundance of *N. pachyderma* (sin.) (Fig. 3). In light of the genotype evidence for the Nordic seas, this Holocene offset in $\delta^{18}\text{O}$ has to be interpreted as an alternation in the species of *N. pachyderma* (dex.) from [Type I (sin.)] during glacial times to a [Type I (dex.)] domination in the interglacial interval. The isotopic offset of about 0.5‰ diminishes in core PS1243 since about 4 kyr ago, and corresponds to a period of increasing abundance of *N. pachyderma* (sin.) and a coeval abundance decrease of *N. pachyderma* (dex.) fairly close to the threshold value of ~5% coiling ratio. This would result in a greater contribution of the right-coiling [Type I (sin.)] genotype to the $\delta^{18}\text{O}$ signal during this time in core PS1243, and explains the reduced offset in $\delta^{18}\text{O}$ between *N. pachyderma* (sin.) and *N. pachyderma* (dex.) compared to that of core M23323. Altogether this strongly supports our conclusion that a species-dependent isotopic shift in *N. pachyderma* (dex.) $\delta^{18}\text{O}$ signal can be inferred

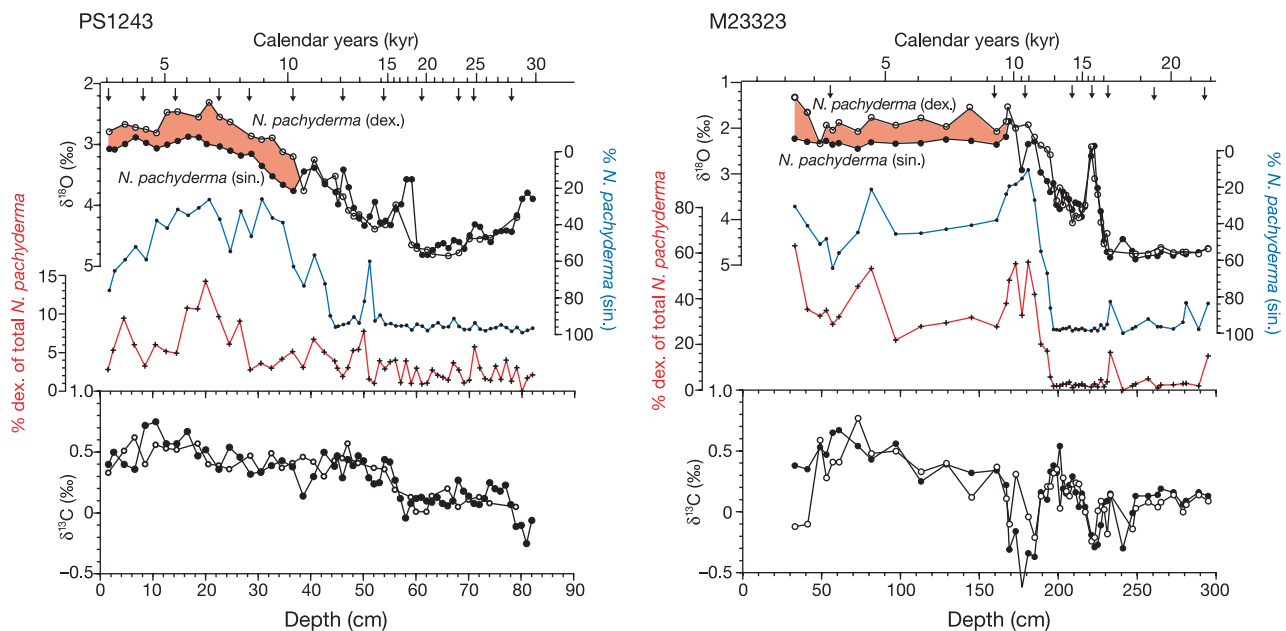


Figure 3 Time series showing climate proxies of two sediment cores versus depth and versus calendar years. The age tie-points of calendar years^{21,29} are indicated on top by arrows. Core P1243 is located close to the Arctic Front (69° N, 7° W). Core M23323 is located under the warm Norwegian Current on the Vøring Plateau (68° N, 6° E) (for core locations see Fig. 1). Top: $\delta^{18}\text{O}$ values of *N. pachyderma* (sin.) (filled circles) and *N. pachyderma* (dex.) (open circles; size fraction 125–250 μm). Middle: coiling ratio

(*N. pachyderma* (dex.) as a percentage of total *N. pachyderma*) shown in red, and relative abundance of *N. pachyderma* (sin.) (given as a percentage of total planktonic species (size fraction >125 μm)) shown in blue; please note reversed y-axis for relative abundance of *N. pachyderma* (sin.). Bottom: $\delta^{13}\text{C}$ values of *N. pachyderma* (sin.) (filled circles) and *N. pachyderma* (dex.) (open circles; size fraction 125–250 μm).

when relative abundances of *N. pachyderma* (dex.) cross a specific threshold value of ~5%.

The $\delta^{18}\text{O}$ isotopic difference of 0.5‰ between the right-coiling forms of *N. pachyderma* [Type I (sin.)] and *N. pachyderma* [Type I (dex.)] in the Holocene translates into a temperature or salinity difference of approximately 2 °C or 2 practical salinity units, respectively. This isotopic difference is an additional increment, superimposed upon the global LGM/Holocene isotopic signature. For all future palaeoceanographic $\delta^{18}\text{O}$ isotope studies carried out on *N. pachyderma* (dex.) at low percentage coiling ratios in the Northern Hemisphere, it will be of considerable importance to account for this derived species dependent isotopic shift of up to ~0.5‰. Interestingly, while the $\delta^{18}\text{O}$ values reveal such a marked difference during the Holocene, there is no systematic offset in $\delta^{13}\text{C}$ between *N. pachyderma* (sin.) and *N. pachyderma* (dex.) over the entire time interval (Fig. 3). Since $\delta^{13}\text{C}$ values of foraminifers are thought to record specific environmental conditions^{23,24}, the Holocene $\delta^{18}\text{O}$ offset may represent a genotype-specific vital effect. However, we have shown that *N. pachyderma* [Type I (sin.)] and *N. pachyderma* [Type I (dex.)] are clearly adapted to different hydrographic regimes, and it remains to be determined whether the species-dependent $\delta^{18}\text{O}$ offset represents annual, seasonal or water-depth hydrological differences, a vital effect or a combination of them. Although our findings add yet another aspect to the interpretation of $\delta^{18}\text{O}$ values of planktonic foraminifers in the Northern Hemisphere, by using the indicator proxy of the coiling ratio, the correct interpretation can be deduced. Clearly the relevance of coiling ratios for isotope records of *N. pachyderma* (dex.) in other palaeoceanographically important regions needs to be addressed.

There is accumulating evidence that several morphospecies do not represent genetically continuous species over their entire environmental adaptive range, as newly discovered potential 'cryptic species' seem to have different environmental preferences^{6–8,25}. Isotopic differences may also be associated with other genotypes of morphospecies commonly used in palaeoceanographic reconstructions^{26,27}. An understanding of the nature of these adaptations will considerably enhance other palaeoceanographic and palaeoclimate proxies if the 'cryptic species' can be distinguished in the fossil record, as was possible for this study. □

Methods

Foraminiferal counts

In down-core sediment samples about 300 individuals of the size fraction >125 µm have been counted, while in surface sediment samples individuals >150 µm have been counted¹⁰. Comparison of counting results of these different size classes in the Nordic seas has shown that there is no significant difference in relative faunal abundances and thereby also coiling ratios²⁸. In the *N. pachyderma* (sin.) dominated polar regions, typical coiling ratios of *N. pachyderma* (dex.) are between 1% and 5 ± 1% from counts of 300 individuals¹⁰. The upper value of 5% coiling ratio has been adopted as the threshold value for the start of the influx of *N. pachyderma* [Type I (dex.)] into the *N. pachyderma* [Type I (sin.)] fossil assemblage. This threshold conforms to the upper values observed in the polar-region core tops (Fig. 1a), the glacial down-core sediments (Fig. 3), and is also approximately the threshold observed in core PS1243 at which the isotopic offset is observed to decrease in the later Holocene (Fig. 3).

Isotope analysis of calcite shells

About 20–30 individuals were picked from the 125–250 µm size fraction, and analysed at the Leibniz Laboratory of Kiel University using a fully automated Kiel Device plus MAT 251 isotope mass spectrometer system. Results are reported in the usual δ -notation relative to PDB, and have an accuracy of 0.07‰ and 0.04‰ for $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$, respectively.

Sampling localities

Foraminifera were collected on board RV *Polarstern* (ARKXV/1 + 2); specimens were obtained using multinet (upper 500 m) at 75° N (from 13° W to 13° E), and from a mixture of multinet and surface pumped water samples in the Nordic seas between 80° N and 60° N.

Isolation and sequencing of SSU gene fragments

DNA extraction, amplification by polymerase chain reaction (PCR) and automated sequencing of either a ~500-base-pair (bp) or a ~1,000-bp region of the terminal 3' end of the foraminiferal SSU rRNA gene was as described previously^{7,8}. The gene fragments for

N. pachyderma (dex.) were directly sequenced. Some limited degree of ambiguity was detected in the gene repeats for *N. pachyderma* (sin.), and consensus sequences were obtained by cloning using a pCR 2.1 TOPO TA cloning kit (Invitrogen) and then sequencing using universal primers.

Received 6 March; accepted 20 May 2003; doi:10.1038/nature01778.

- Ericson, D. B. Coiling direction of *Globigerina pachyderma* as a climatic index. *Science* **130**, 219–220 (1959).
- CLIMAP Project Members The surface of the ice-age Earth. *Science* **191**, 1131–1137 (1976).
- Bond, G. C., Broecker, W., Johnsen, S. & McManus, J. Correlations between climate records from North Atlantic sediments and Greenland ice. *Nature* **365**, 143–147 (1993).
- Bé, A. W. & Tolderlund, D. S. in *The Micropaleontology of the Oceans* (eds Funnel, B. M. & Riedel, W. R.) 105–149 (Cambridge Univ. Press, Cambridge, 1971).
- Bandy, O. L. Origin and development of *Globorotalia* (*Turborotalia*) *pachyderma* (Ehrenberg). *Micropaleontology* **18**, 294–318 (1972).
- de Vargas, C., Norris, R., Zaninetti, L., Gibb, S. W. & Pawlowski, J. Molecular evidence of cryptic speciation in planktonic foraminifers and their relation to oceanic provinces. *Proc. Natl Acad. Sci. USA* **96**, 2864–2868 (1999).
- Darling, K. F. et al. Molecular evidence for genetic mixing of Arctic and Antarctic subpolar populations of planktonic foraminifers. *Nature* **404**, 43–47 (2000).
- Stewart, I. A., Darling, K. F., Kroon, D., Wade, C. M. & Troelstra, S. R. Genotypic variability in subarctic Atlantic planktic foraminifera. *Mar. Micropaleontol.* **43**, 143–153 (2001).
- Kipp, N. G. New transfer function for estimating past sea-surface conditions from sea-bed distribution of planktonic foraminiferal assemblages in the North Atlantic. *Geol. Soc. Mem.* **145**, 3–41 (1976).
- Pflaumann, U., Duprat, J., Pujol, C. & Labeyrie, L. D. SIMMAX: A modern analog technique to deduce Atlantic sea surface temperatures from planktonic foraminifera in deep-sea sediments. *Paleoceanography* **11**, 15–35 (1996).
- Healy-Williams, N., Williams, D. F. & Ehrlich, R. Quantifications of morphotypes in *Neogloboquadrina pachyderma* using Fourier shape analysis. *Antarct. J. US* **18**, 138–140 (1984).
- Williams, D. F., Ehrlich, R., Spero, H. J., Healy-Williams, N. & Gary, A. C. Shape and isotopic differences between conspecific foraminiferal morphotypes and resolution of paleoceanographic results. *Paleoceanogr. Palaeoclimatol. Palaeoecol.* **64**, 153–162 (1988).
- Cifelli, R. Observations of *Globigerina pachyderma* (Ehrenberg) and *Globigerina incompacta* Cifelli from the North Atlantic Ocean. *J. Foram. Res.* **3**, 157–166 (1973).
- Huber, R., Meggers, H., Baumann, K.-H., Raymo, M. E. & Henrich, R. Shell size variation of the planktonic foraminifer *Neogloboquadrina pachyderma* sin. in the Norwegian Greenland Sea during the last 1.3 Myrs: Implications for paleoceanographic reconstructions. *Paleoceanogr. Palaeoclimatol. Palaeoecol.* **160**, 193–212 (2000).
- Kellogg, T., Duplessy, J. & Shackleton, N. Planktonic foraminiferal and oxygen isotopic stratigraphy and paleoclimatology of Norwegian Sea deep-sea cores. *Boreas* **7**, 61–73 (1978).
- Jansen, E. The use of stable oxygen and carbon isotope stratigraphy as a dating tool. *Quat. Int.* **1**, 151–166 (1989).
- Bauch, H. A. in *Land-ocean Systems in the Siberian Arctic: Dynamics and History* (eds Kassens, H. et al.) 601–613 (Springer, Berlin, 1999).
- Oppo, D. W., McManus, J. F. & Cullen, J. L. Abrupt climate events 500,000 to 340,000 years ago: Evidence from subpolar North Atlantic sediments. *Science* **279**, 1335–1338 (1998).
- McManus, J., Oppo, D. & Cullen, J. A 0.5-million-year record of millennial-scale climate variability in the North Atlantic. *Science* **283**, 971–975 (1999).
- Oppo, D., Keigwin, L. D., McManus, J. F. & Cullen, J. L. Persistent suborbital climate variability in marine isotope stage 5 and Termination II. *Paleoceanography* **16**, 280–292 (2001).
- Simstich, J. Die ozeanische Deckschicht des Europäischen Nordmeers im Abbild stabiler Isotope von Kalkgehäusen unterschiedlicher Planktonforaminiferenarten. *Ber.-Rep. Inst. Geowiss., Univ. Kiel* **2**, 1–96 (1999).
- Emiliani, C. Pleistocene temperatures. *J. Geol.* **63**, 538–578 (1955).
- Spero, H. J., Bijma, J., Bemis, B. & Lea, D. Effects of sea water carbonate chemistry on planktonic foraminifera carbon and oxygen isotope values. *Nature* **390**, 497–500 (1997).
- Bauch, D., Erlenkeuser, H., Winckler, G., Pavlova, G. & Thiede, J. Carbon isotopes and habitat of polar planktic foraminifera in the Okhotsk Sea: The "Carbonate Ion Effect" under natural conditions. *Mar. Micropaleontol.* **45**, 83–99 (2002).
- de Vargas, C., Renaud, S., Hilbrecht, H. & Pawlowski, J. Pleistocene adaptive radiation in *Globorotalia truncatulinoides*: Genetic, morphologic, and environmental evidence. *Paleobiology* **27**, 104–125 (2001).
- Bemis, B. E., Spero, H. J. & Thunell, R. C. Using species-specific paleotemperature equations with foraminifera: A case study in the Southern California Bight. *Mar. Micropaleontol.* **46**, 405–430 (2002).
- Darling, K. F., Kucera, M., Wade, C. M., von Langen, P. & Pak, D. Seasonal occurrence of genetic types of planktonic foraminiferal morphospecies in the Santa Barbara Channel. *Paleoceanography* **18**, 1032, doi: 10.1029/2001PA000723 (2003).
- Kandiano, E. S. & Bauch, H. A. Implications of planktic foraminiferal size fractions for the glacial-interglacial paleoceanography of the polar North Atlantic. *J. Foram. Res.* **32**, 245–251 (2002).
- Bauch, H. A. et al. A multiproxy reconstruction of the evolution of deep and surface waters in the subarctic Nordic seas over the last 30,000 years. *Quat. Sci. Rev.* **20**, 659–678 (2001).

Acknowledgements We thank the crew and scientists of RV *Polarstern* (ARK XV) for their efforts. Sampling on board was conducted by J. Netzer and E. Stangeew. Part of this work was originally conducted within SFB313, and we thank J. Rumohr for unpublished data. D.B. was supported by the Deutsche Forschungsgemeinschaft. This work was supported by the NERC and the Leverhulme Trust.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.B. (dbauch@geomar.de). The sequences determined in this study are available at GenBank under accession numbers AF250117 [Type I (dex.)] and AY305329 [Type I (sin.)].

Rapid evolution drives ecological dynamics in a predator–prey system

Takehito Yoshida*, Laura E. Jones*, Stephen P. Ellner*, Gregor F. Fussmann*† & Nelson G. Hairston Jr*

* Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA

Ecological and evolutionary dynamics can occur on similar timescales^{1–7}. However, theoretical predictions of how rapid evolution can affect ecological dynamics⁸ are inconclusive and often depend on untested model assumptions⁸. Here we report that rapid prey evolution in response to oscillating predator density affects predator–prey (rotifer–algal) cycles in laboratory microcosms. Our experiments tested explicit predictions from a model for our system that allows prey evolution⁹. We verified the predicted existence of an evolutionary tradeoff between algal competitive ability and defence against consumption, and examined its effects on cycle dynamics by manipulating the evolutionary potential of the prey population. Single-clone algal cultures (lacking genetic variability) produced short cycle periods and typical quarter-period phase lags between prey and predator densities, whereas multi-clonal (genetically variable) algal cultures produced long cycles with prey and predator densities nearly out of phase, exactly as predicted. These results confirm that prey evolution can substantially alter predator–prey dynamics, and therefore that attempts to understand population oscillations in nature^{10,11} cannot neglect potential effects from ongoing rapid evolution.

In our experiments, parthenogenetic planktonic rotifers, *Brachionus calyciflorus*, consumed unicellular, obligately asexual¹² green algae, *Chlorella vulgaris*. These were cultured in continuous flow-through chemostat systems in a nitrogen-limited medium. A simple differential-equation model predicts the transitions between the qualitatively different population dynamics observed in this system (equilibria, population cycles and extinction) as a function of the nutrient concentration of the medium and the dilution (flow-through) rate of the chemostat¹³. However, cycle periods were far longer than predicted, and observed predator and prey cycles were almost exactly out of phase (predator maxima and prey minima were nearly simultaneous, and vice versa), which is impossible in a conventional predator–prey model¹³. These unexpected cycles had extended periods when algal biomass was high but rotifer densities remained low, followed by increased rotifer growth even though algal densities remained nearly constant.

To identify possible mechanisms that might account for these observations, we developed simple mathematical models for a series of alternative biologically plausible hypotheses⁹: changes in algal carbon:nitrogen ratio as a function of nutrient availability; rotifer self-limitation through toxin production or decreased egg viability when prey are scarce; and rapid prey evolution resulting from an evolutionary tradeoff between algal competitive ability and defence against rotifer predation. The model with rapid prey evolution reproduced the observed cycles much more accurately than our original model, whereas the non-evolutionary models failed to produce cycles similar to those observed⁹.

Here we present experimental tests of the prediction that the qualitative properties of rotifer–algal cycles in our system are the result of rapid prey evolution. To test for the evolutionary tradeoff postulated by the prey evolution model, we manipulated algal growth conditions. Algae cultivated under constant and intense

rotifer grazing pressure became lower in food value (in the sense that rotifers that fed on these cells had a lower population growth rate) and were heritably smaller and competitively inferior relative to algae grown in the absence of rotifers but with a comparable mortality rate imposed by an elevated chemostat dilution rate (Table 1). These results show the existence of different genotypes in the algal population, and the presence of a tradeoff between algal food value and competitive ability among clones. The heritable response to rotifer predation also shows that the selected low-food-value algal genotypes are better able to survive rotifer grazing, although the mechanism for this remains to be determined.

We added genetic structure, and therefore the possibility for algal evolution, to our original model¹³ by representing the algal population as an assortment of clones differing in food value and competitive ability (Fig. 1). Clonal selection is appropriate because *C. vulgaris* reproduces asexually¹² and because we showed (above) the presence of different genotypes—clones—in our laboratory population (Table 1). The model incorporates the observed evolutionary tradeoff by assuming that lower-food-value clones have a higher half-saturation constant for nutrient uptake (Fig. 1a, c; see Methods for additional details).

A fundamental prediction of the multiple-clone model is that prey evolution can substantially alter the population cycles (Fig. 1). Model simulations for systems composed of randomly generated sets of one, two, three, five or seven clones show that when multiple clones are present, long cycles are possible (Fig. 1b, d) irrespective of the tradeoff curve specified between food value and competitive ability (Fig. 1a, c). Cycle lengths for two or more coexisting clones comprised a range of periods. Short cycles can occur when the algal population consists of two very similar clones, but with more than two clones short cycles become highly unlikely. In contrast, a single-clone system always produces relatively short cycles, regardless of that clone's phenotype. Moreover, population cycles in a single-clone system exhibit the classic predator–prey phase relations in which the peaks in predator abundance follow the prey peaks by one-quarter of a cycle (Fig. 1e). However, if the prey population consists of two or more phenotypically divergent clones, and predation by rotifers in combination with competition for nutrients results in rapid changes in algal genotype frequencies, the resulting multi-clone population cycles show phase relations like those observed, with algae and rotifers almost exactly out of phase (Fig. 1f). These properties result from the increasing dominance of low-food-value clones as the algal population is grazed down. Therefore, when algal density increases again after the rotifer population has crashed, the rotifer population does not immediately respond. Rotifers cannot increase until the algae return to high density, at which point higher-food-value clones (which are better competitors) increase in abundance. Rotifer grazing then intensifies and the cycle begins anew.

To test these predictions we manipulated prey evolution by altering clonal diversity (that is, genetic variability) in the prey population. We initiated replicated chemostat trials either with a single clone as an evolutionarily 'stagnant' population or with multiple clones comprising an evolutionarily 'active' one, at dilution rates and nutrient concentrations that had previously been shown to produce population cycles¹³. We then compared cycle periods and phase relations between the treatments.

Our results are strikingly consistent with theoretical predictions.

Table 1 Evolutionary tradeoff for *C. vulgaris*

Treatment	Algal food value (d ^{−1})	Competitive ability (d ^{−1})
Rotifers present	1.07 ± 0.05	1.25 ± 0.02
Rotifers absent	1.58 ± 0.05	1.73 ± 0.03

Values are means ± s.e.m. Algal food value was measured by rotifer population growth rate when feeding on algae at sufficiently high concentration (see Methods). Competitive ability was estimated by algal growth rate under nutrient-deficient conditions (see Methods).

† Present address: Institut für Biochemie und Biologie, Universität Potsdam, Maulbeerallee 2, D-14469 Potsdam, Germany.

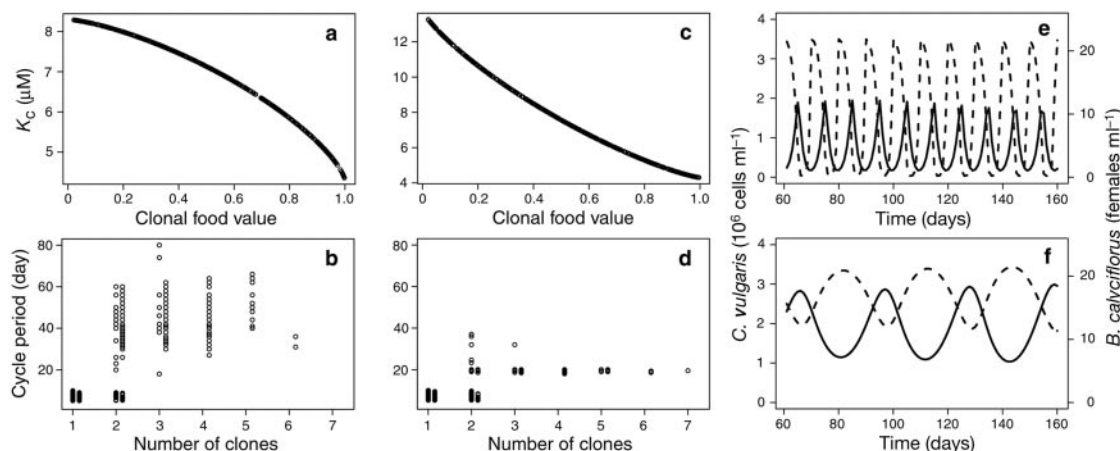


Figure 1 Predicted effects of prey clonal diversity on population dynamics. **a, c**, Modelled tradeoff curves between algal food value and competitive ability (represented by the half-saturation constant, K_c , for nutrient uptake by algae), with $\alpha' > 1$ (**a**) and $\alpha' < 1$ (**c**). **b, d**, Effects of clonal diversity on predator–prey cycle length for the two tradeoff curves between algal food value and competitive ability, with $\alpha' > 1$ (**b**) and $\alpha' < 1$ (**d**). Cycle periods in **b** and **d** are the result of 700 simulation runs each for one, two, three, five or seven clones present at the start of the run. Cycle periods are plotted against the number

of clones remaining after running the model long enough for competitively inferior clones to be eliminated. 'Degenerate' cases (in which some of the initially present clones were eliminated) are plotted just to the right of non-degenerate cases in **b** and **d**. **e, f**, Cycles predicted by the model for the *Brachionus* predator (solid line) and *Chlorella* prey (dashed line) in a single-clone system (**e**) and a multiple-clone system (**f**). Model equations and further technical details are given in the Supplementary Information.

Population cycles of the single-clone evolutionarily stagnant treatment were much shorter and of smaller variance than those of the multiple-clone treatment (Figs 2 and 3). In addition, the single-clone treatments exhibited 'classic' predator–prey cycles with roughly a quarter-cycle delay between prey and predator maxima

(Fig. 2; mean phase lag as a fraction of cycle period, 32%; range 26–36%). By contrast, in the multiple-clone evolutionarily active treatment, the predator and prey maxima were almost exactly out of phase (Fig. 2), as predicted by the clonal evolutionary model (mean phase lag, 56%; range 41–68%). The clonal evolutionary model is

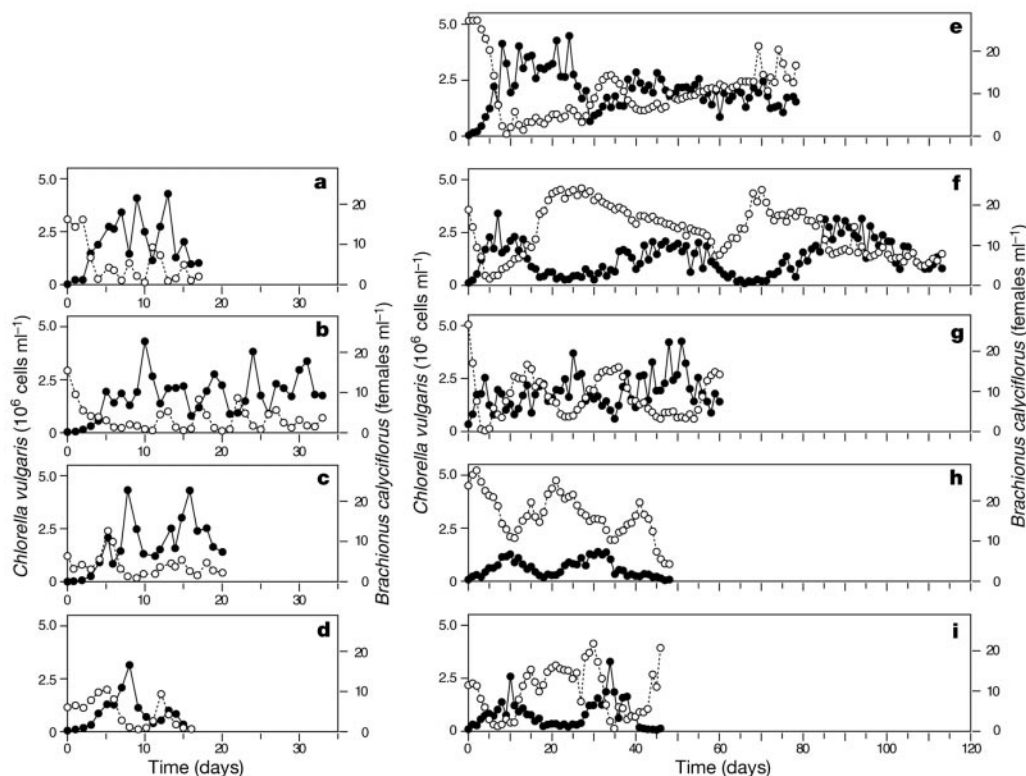


Figure 2 Experimental results showing the population cycles of rotifer–algae systems. **a–d**, Single-clone algal populations; **e–i**, multiple-clone algal populations. Filled circles, *B. calyciflorus* (predator); open circles, *C. vulgaris* (prey). In **a–d** there are short-period predator–prey cycles with the classical phase relations; in **e–i** there are long cycles with predator and prey oscillations nearly out of phase with each other. Dilution rates, δ (d^{-1}) **a**, 0.57; **b**, 0.65; **c**, 0.67; **d**, 0.68; **e**, 0.72; **f**, 0.64; **g**, 0.69; **h**, 0.95; **i**, 1.00. Data for four

(**f–i**) of the five multiple-clone trials were obtained from ref. 13. The fifth trial is a new experiment, conducted in parallel with the single-clone trials, to verify that the lapse of time did not result in any change of our rotifer or algal stock cultures that would change the qualitative dynamics of a system with multiple algal clones. Note that population cycles at relatively high dilution rates (**h, i**) were also long and with out-of-phase relations, as our mathematical model predicts.

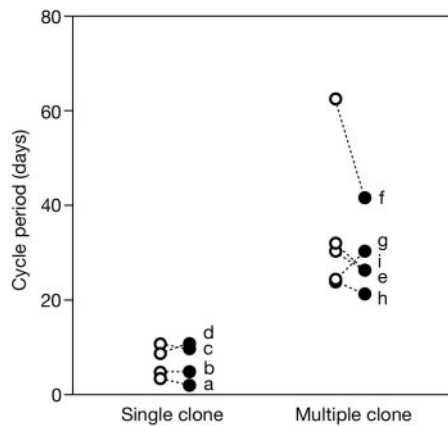


Figure 3 Cycle periods for experimental populations of *B. calyciflorus* (filled circles) and *C. vulgaris* (open circles). Lines connect period estimates from the same experimental trial, and the labelling of points corresponds to the panel labels in Fig. 2. The large difference between prey and predator period estimates by spectral analysis in experiment f resulted from the strongly non-sinusoidal shape of the cycles.

also able to match the experimentally determined bifurcation diagram of the system¹³ (transition points between stable and cyclic dynamics as the dilution rate is varied) more accurately than the non-evolutionary model (see Supplementary Information).

Predators can also evolve rapidly⁷. Although we have not yet investigated predator evolution, it is significant that prey evolution alone brought our model into close agreement with experiments. Two explanations are possible. One is that the selection pressure of predators on prey might be stronger than that of prey on predators (the 'life-dinner' dichotomy^{14,15}). Alternatively, predator evolution simply might not have much influence on our system's dynamics. Although inducible defences by prey are often observed in nature¹⁶, none are evident in our system. Grazing-adapted cells are smaller but otherwise identical to non-adapted cells in morphology and chemical composition (C:N ratio; see Methods), and both types were predominantly unicellular. Inducible defences can also be ruled out as the cause of long-period cycles, because the cycle length and phase relations are both consequences of the fact that algal food value remains poor even after several algal generations without rotifer predation⁹.

Our results provide experimental support for model-based approaches to identifying the causes of natural population cycles^{11,17}. An ideal experimental test removes or alters a candidate causal process (as we have done here) and compares observed with predicted effects on dynamics. However, in natural systems such tests are typically infeasible because of the large spatial and temporal scales at which systems would have to be manipulated¹⁸. Even when such experiments are possible, their interpretation can be problematic¹¹. An alternative is to construct mechanistic models corresponding to each potential causal mechanism and to compare each model's consistency with available observational and experimental data^{11,17,19–21}. Our results here confirm the conclusion previously derived by this approach: that among all candidate causal processes only rapid prey evolution could account for our system's dynamic patterns.

We have shown that ongoing rapid prey evolution can alter population dynamics (cycle period and phase relations) in a live predator–prey system. Population cycles can lead to fluctuating selection, so that rapid prey evolution occurs indefinitely and reciprocally influences population dynamics. The interactions governing ecological dynamics are then continually changing through rapid evolution. Our study indicates that ecologists must consider ongoing rapid evolution when exploring the mechanisms under-

lying the dynamics of populations and food webs. Evolutionary and ecological dynamics must be understood in concert. □

Methods

Selection experiment

To assess variation in food value and competitive ability, *C. vulgaris* populations were exposed in two chemostats either to constant intense rotifer predation (dilution rate $\delta = 0.19 \text{ d}^{-1}$) or to unselective mortality from elevated washout ($\delta = 1.24 \text{ d}^{-1}$). Algal C:N ratios in the two treatments were similar (molar C:N = 13–14) and lower than in algae cultured under nutrient-deficient conditions ($\delta = 0.16 \text{ d}^{-1}$, molar C:N = 20), indicating that both treatments experienced an equivalent high availability of nutrients. Algae were harvested after each chemostat had reached equilibrium and then used to determine the effects of previous selection under common conditions.

Algal food value was determined by bioassay: rotifer population growth rate when fed with algae at a density ($7 \times 10^6 \text{ cells ml}^{-1}$) above the incipient limiting concentration²². Experimental procedures followed ref. 23. Algal competitive ability was estimated by measuring algal population growth rate under nutrient-deficient conditions in short-term batch cultures: 100 ml culture medium with $1 \mu\text{M}$ nitrate. Algae were inoculated at $500 \text{ cells ml}^{-1}$, grown at 25°C under constant $120 \mu\text{E m}^{-2} \text{ s}^{-1}$ illumination (as in the chemostat experiments), and cell density and size distribution were monitored daily. Algal populations grew exponentially for the first 4–5 days; maximum growth rate was determined as the maximum derivative of a local polynomial curve fitted to the logarithm of algal density against time.

Simulation analysis

Model equations, parameter values and details are given in the Supplementary Information. Our model is based on that of ref. 13, a system of nonlinear differential equations for the state variables: limiting nutrient, algal abundance, breeding and total rotifer abundances. The two rotifer variables represent rotifer age structure, with rotifers gradually ceasing to breed with age; in other respects it is a standard chemostat model with Monod equations²⁴ for algal nutrient uptake, and for predation by rotifers on algae, with fecundity proportional to consumption. Here, we replaced the single algal variable by a series of algal clones. Clones are characterized by their relative food value to rotifers, p , ranging between 1 ('good') and ~ 0 ('poor'), with lower p resulting in reduced predation risk. The defensive 'low-food-value' trait comes at the expense of reduced competitive ability (Table 1). To model this tradeoff, lower- p clones are given a higher half-saturation constant for nutrient uptake and are therefore poorer competitors when nutrients are scarce (our previous model⁹ made different assumptions that are inconsistent with subsequent experimental results (Table 1, and T.Y., unpublished data) and used quantitative trait dynamics rather than clonal selection).

We initialized simulations with $q = 1, 2, 3, 5$ or 7 clones with random p values (uniformly distributed on $[0.02, 1]$) and competitive ability assigned according to one of two different tradeoff curves (Fig. 1a, c). Competitive ability (represented by the half-saturation constant K_c) is related to food value by $K_c = K + \alpha(1 - p^\alpha)^{1/\alpha'}$, where K is the minimum value applying when $p = 1$, α is a scale parameter describing the cost of decreased food value and α' is a shape parameter determining the concavity of the tradeoff. Estimation of tradeoff curves is described in Supplementary Table 1. Clone sets were randomly drawn 700 times for each q . Clones comprising less than 1% of the final algal population were considered extinct. We then recorded the number of surviving clones and the cycle period (scored as period = 0 for coexistence at equilibrium).

Chemostat experiments

We established stock cultures of *Chlorella vulgaris* (UTEX no. 26) and *Brachionus calyciflorus* (taken originally from Milwaukee harbour, Wisconsin, and provided by M. Boraas). Although we do not know how much clonal (genotypic) variation for food value and competitive ability was present in our algal stock culture, populations derived from it in our selection experiment have been shown to possess distinct heritable phenotypes (T.Y., unpublished data), so we conclude that it consists of multiple clones. Isolated single algal clones were established by spot-plating cells from the stock culture on agar plates. We chose one of these arbitrarily for use in the experiments reported here.

The experimental chemostats followed ref. 13. To prevent algal contamination in the treatments using single algal clones, rotifers were taken from stock culture, rinsed thoroughly with sterilized medium, and held in sterilized medium for at least 6 h to ensure that all algal cells in their stomachs were digested. Rotifers were then thoroughly rinsed again with sterilized medium before being used to initiate chemostat trials. We set the dilution rate ($0.57\text{--}1.00 \text{ d}^{-1}$) and nutrient concentration ($80 \mu\text{M}$ nitrate) to give population cycles (ref. 13 and our model). Organisms were sampled daily through ports near the bottom and top of each chemostat. Rotifers were counted under a dissecting microscope and algae were counted with a particle counter (CASY 1; Schärfe, Reutlingen, Germany). Organism abundance data are means of duplicate samples. We determined cycle periods by spectral analysis of organism counts after rotifer populations were established (more than one female ml^{-1}) using spec.pgram in the R language²⁵, and phase lags from the cross-correlation function between predator and prey counts (ccf in R).

Received 28 February; accepted 28 April 2003; doi:10.1038/nature01767.

- Hairston, N. G. Jr *et al.* Rapid evolution revealed by dormant eggs. *Nature* **401**, 446 (1999).
- Huey, R. B., Gilchrist, G. W., Carlson, M. L., Berrigan, D. & Serra, L. Rapid evolution of a geographic cline in size in an introduced fly. *Science* **287**, 308–309 (1999).
- Hendry, A. P., Wenbin, J. K., Bentzen, P., Volk, E. C. & Quinn, T. P. Rapid evolution of reproductive isolation in the wild: Evidence from introduced salmon. *Science* **290**, 516–518 (2000).
- Thompson, J. N. *et al.* Frontiers of ecology. *BioScience* **51**, 15–24 (2001).

5. Palumbi, S. R. *The Evolution Explosion: How Humans Cause Rapid Evolutionary Change* (W. W. Norton, New York, 2001).
6. Sinervo, B., Svensson, E. & Comendant, T. Density cycles and an offspring quantity and quality game driven by natural selection. *Nature* **406**, 985–988 (2000).
7. Bohannan, B. J. M. & Lenski, R. E. Linking genetic change to community evolution: Insights from studies of bacteria and bacteriophage. *Ecol. Lett.* **3**, 362–377 (2000).
8. Abrams, P. A. The evolution of predator–prey interactions: theory and evidence. *Annu. Rev. Ecol. Syst.* **31**, 79–105 (2000).
9. Shertzer, K. W., Ellner, S. P., Fussmann, G. F. & Hairston, N. G. Jr Predator–prey cycles in an aquatic microcosm: Testing hypotheses of mechanism. *J. Anim. Ecol.* **71**, 802–815 (2002).
10. Berryman, A. (ed.) *Population Cycles: The Case for Trophic Interactions* (Oxford Univ. Press, 2002).
11. Turchin, P. *Complex Population Dynamics: A Theoretical/Empirical Synthesis* (Princeton Univ. Press, 2003).
12. Pickett-Heaps, J. D. *Green Algae: Structure, Reproduction and Evolution in Selected Genera* (Sinauer Associates, Sunderland, Massachusetts, 1975).
13. Fussmann, G. F., Ellner, S. P., Shertzer, K. W. & Hairston, N. G. Jr Crossing the Hopf bifurcation in a live predator–prey system. *Science* **290**, 1358–1360 (2000).
14. Vermeij, G. J. *Evolution and Escalation: An Ecological History of Life* (Princeton Univ. Press, 1987).
15. Vermeij, G. J. The evolutionary interaction among species: selection, escalation, and coevolution. *Annu. Rev. Ecol. Syst.* **25**, 219–236 (1994).
16. Tollrian, R. & Harvell, C. D. (eds) *The Ecology and Evolution of Inducible Defenses* (Princeton Univ. Press, 1999).
17. Kendall, B. E. *et al.* Why do populations cycle? A synthesis of statistical and mechanistic modeling approaches. *Ecology* **80**, 1789–1805 (1999).
18. Lambin, X., Krebs, C. J., Moss, R. & Yoccoz, N. G. in *Population Cycles: The Case for Trophic Interactions* (ed. Berryman, A.) 155–176 (Oxford Univ. Press, 2002).
19. Hillborn, R. & Mangel, M. *The Ecological Detective: Confronting Models with Data* (Princeton Univ. Press, 1997).
20. McCauley, E., Nisbet, R. M., Murdoch, W. W., de Roos, A. M. & Gurney, W. S. C. Large-amplitude cycles of *Daphnia* and its algal prey in enriched environments. *Nature* **402**, 653–656 (1999).
21. Turchin, P. *et al.* Dynamical effects of plant quality and parasitism on population cycles of larch budmoth. *Ecology* (in the press).
22. Halbach, U. & Halbach-Keup, G. Quantitative relations between phytoplankton and the population dynamics of the rotifer *Brachionus calyciflorus* Pallas. Results of laboratory experiments and field studies. *Arch. Hydrobiol.* **73**, 273–309 (1974).
23. Rothhaupt, K. O. Algal nutrient limitation affects rotifer growth rate but not ingestion rate. *Limnol. Oceanogr.* **40**, 1201–1208 (1995).
24. Monod, J. La technique de culture continue: theorie et applications. *Ann. Inst. Pasteur Lille* **79**, 390–410 (1950).
25. Ihaka, R. & Gentleman, R. R. a language for data analysis and graphics. *J. Comp. Graph. Stat.* **5**, 299–314 (1996).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank B. Kendall, K. Shertzer, J. Urabe and members of the EEB theoretical ecology 'lunch-bunch' for comments on the manuscript; A. Sasaki and C. Aquadro for discussions on clonal evolution; and M. Armsby, S. Hammer, M. Hung, C. Kearns, K. Keller and J. Meyer for assistance with the experiments. The study was supported by a grant from the Andrew W. Mellon Foundation to S.P.E. and N.G.H.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.G.H. (NGH1@cornell.edu).

Complex hybrid origin of genetic caste determination in harvester ants

Sara Helms Cahan & Laurent Keller

Institute of Ecology, University of Lausanne, CH-1015 Lausanne, Switzerland

Caste differentiation and division of labour are the hallmarks of insect societies¹ and at the root of their ecological success². Kin selection predicts that caste determination should result from environmentally induced differences in gene expression^{3,4}, a prediction largely supported by empirical data⁵. However, two exceptional cases of genetically determined caste differentiation have recently been found in harvester ants^{6–8}. Here we show that genetic caste determination evolved in these populations after complex hybridization events. We identified four distinct genetic lineages, each consisting of unique blends of the genomes of

the parental species, presumably *Pogonomyrmex barbatus* and *P. rugosus*. Crosses between lineages H1 and H2 and between J1 and J2 give rise to workers, whereas queens develop from within-lineage matings. Although historical gene flow is evident, genetic exchange among lineages and between lineages and the parental species no longer occurs. This unusual system of caste determination seems to be evolutionarily stable.

Pogonomyrmex barbatus and *P. rugosus* are common harvester ant species whose ranges broadly overlap in southwestern North America⁹. In both species, a population within the overlap zone in southwestern New Mexico has been found to possess a system of genetic caste determination^{6–8}. By contrast, caste determination in populations outside the overlap zone is non-genetic^{7,8}, as is typical in ants¹⁰.

To gain insight into the origin of genetic caste determination and the relationship between its occurrences in the two species, we conducted a genetic study of the two adjacent sites (Hidalgo⁶ and Junction⁷) in which genetic caste determination has been described, and also in allopatric populations of *P. rugosus* and *P. barbatus* (see Methods). The three classes of genetic markers (allozymes, microsatellites and mitochondrial sequence data) revealed that the two sites are each composed of a distinct pair of interbreeding lineages, H1 (red-male⁶) and H2 (black-male⁶) at Hidalgo, and J1 (lineage X (ref. 7)) and J2 (lineage 4 (ref. 7)) at Junction. Each lineage has a unique multilocus genotype (Table 1) and is strongly differentiated from all other lineages (Nei's *D*, range = 0.53–1.50). Each lineage also contains a diagnostic monophyletic set of mitochondrial haplotypes (Fig. 1), showing a lack of genetic exchange across lineages.

As in previous studies^{6–8}, there were marked differences in the genomic composition of queens and workers at both Hidalgo and Junction. Of the 42 winged (young) queens collected at Hidalgo, 40 contained either an H1–H1 or H2–H2 genome, the remaining two having H1–H2 genomes. All 40 workers had an H1–H2 genome. A similar pattern was uncovered at Junction. Of the 38 winged queens, 37 had a J1–J1 or J2–J2 genome while one queen and all 35 workers had J1–J2 genomes. The few inter-lineage winged queens produced seem to have low reproductive success; we found no colonies displaying genotypes consistent with an inter-lineage mother (queen) at these sites or at five other sites surveyed in the region (22–40 colonies sampled per site, S. Helms Cahan and L. Keller, unpublished observations). We also did not find a single individual with an H–J genome of any type (H1–J1, H1–J2, H2–J1 or H2–J2), indicating that crosses between lineages from the two sites either do not occur or fail to give rise to viable females.

In addition to being genetically isolated from one another, all four lineages are genetically distinct from the *P. rugosus* and *P. barbatus* populations with non-genetic caste determination. Genetic distances between each species and the four lineages were uniformly high (*P. rugosus*, 0.35–0.85; *P. barbatus*, 0.56–1.39), with diagnostic differences at one or more nuclear loci between the two species and each of the four lineages (see Supplementary Information). Haplotypes of all four lineages were also clearly differentiated from both *P. rugosus* and *P. barbatus* (Fig. 1). Thus, neither *P. barbatus* nor *P. rugosus* seems to be currently linked by gene flow with any of the four lineages.

Two general hypotheses have been proposed for the origin of a two-lineage genetic caste system. The first is that a heterozygosity-based caste locus evolved within species, resulting in the splitting of the ancestral gene pool into two diverging lineages⁷. The second is that the evolution of genetic caste determination is associated with interspecific hybridization^{6,8}. Our results show that the presence of distinct lineages within populations resulted from hybridization, most probably between *P. rugosus* and *P. barbatus*. Across nuclear markers, the two interbreeding lineages at each site clustered with different parental species: H1 and J1 with *P. rugosus*, and H2 and J2 with *P. barbatus* (Fig. 2). Moreover, the *cox1* mitochondrial haplotypes of J1 and J2 grouped together are paraphyletic (Fig. 1): the J1 clade is most closely related to *P. rugosus*, whereas that of J2 is

Table 1 Allele frequencies of loci showing fixed differences between one or more pairs of lineages

Locus	Allele	H1	H2	J1	J2	Locus	Allele	H1	H2	J1	J2
<i>PGI</i>	1	—	—	0.182	—	<i>Pb5</i>	216	—	0.063	—	0.022
	2	—	—	0.818	—		222	—	—	—	0.022
	3	1.000	—	—	—		224	—	0.063	—	0.065
	4	—	1.000	—	1.000		226	—	0.875	—	—
	<i>n</i>	17	9	11	20		228	—	—	—	0.087
<i>PGM-2</i>	1	1.000	—	1.000	—		232	1.000	—	0.591	0.696
	2	—	1.000	—	1.000		234	—	—	—	0.022
	<i>n</i>	8	4	11	15		236	—	—	—	0.087
							240	—	—	0.045	—
							242	—	—	0.318	—
<i>EST-2</i>	1	—	1.000	1.000	1.000		244	—	—	0.045	—
	2	1.000	—	—	—		<i>n</i>	24	8	11	23
	<i>n</i>	9	4	11	19						
						<i>Pb7</i>	139	—	0.071	—	—
							140	—	0.071	—	—
<i>Myrt3</i>	180	1.000	1.000	0.682	—		143	—	0.786	—	—
	182	—	—	—	1.000		148	—	0.071	0.042	—
	188	—	—	0.227	—		152	0.563	—	0.125	0.024
	190	—	—	0.091	—		154	0.063	—	0.625	0.476
	<i>n</i>	22	8	11	21		156	0.375	—	0.167	—
<i>L18</i>	156	—	—	—	0.023		158	—	—	—	0.024
	158	—	—	—	0.023		160	—	—	0.042	—
	159	—	—	—	0.023		162	—	—	—	0.048
	161	—	—	—	0.182		168	—	—	—	0.190
	163	—	—	—	0.205		172	—	—	—	0.048
	164	0.289	—	—	—		178	—	—	—	0.024
	165	—	0.143	—	0.136		182	—	—	—	0.048
	166	0.158	—	—	—		184	—	—	—	0.095
	167	—	0.429	—	0.091		186	—	—	—	0.024
	168	0.368	—	0.200	—		<i>n</i>	24	7	12	21
	169	—	0.143	—	0.227	<i>Pb8</i>	268	—	—	0.292	—
	170	0.132	—	—	—		270	—	—	0.708	—
	171	—	0.071	—	0.091		276	0.340	—	—	—
	180	—	—	0.200	—		278	0.560	—	—	0.326
	181	—	0.071	0.050	—		280	0.060	—	—	—
	182	—	—	0.150	—		282	0.020	0.786	—	—
	184	0.026	—	0.050	—		284	—	0.071	—	0.065
	185	—	0.071	—	—		286	0.020	—	—	0.261
	186	0.026	—	0.150	—		288	—	—	—	0.087
	187	—	0.071	—	—		290	—	—	—	0.196
	188	—	—	0.150	—		292	—	0.143	—	0.043
	192	—	—	0.050	—	317	—	—	—	0.022	—
	<i>n</i>	19	7	10	22	<i>n</i>	25	7	12	23	

Allele frequencies were based on genotypes of a single winged queen from each colony. Allozyme alleles are numbered consecutively by order of migration on cellulose acetate gels; microsatellite allele numbers refer to allele size as measured on an ABI 377 automated sequencer. An additional three allozyme loci (*PGM-1*, *EST-1*, *HEX*) showed no fixed differences between any pair of lineages. Allele frequencies of the four lineages at these loci, as well as allele frequencies of *P. rugosus* and *P. barbatus* populations at all loci, are given in the Supplementary Information.

associated with *P. barbatus* (Interior Branch Test, branch length for (J1, *P. rugosus*)(J2, *P. barbatus*) = 0.047 ± 0.011 (s.e.m.), confidence probability = 1).

Although lineages have an overall affinity with one of the two parental species, inspection of the genotypic data reveals a surprising amount of genetic mixing. On average, only 63% (range 59–73) of the alleles in each lineage also occur in its putative parent.

However, when the alleles of both parental species are considered the percentage of alleles accounted for increases to 85% (range 71–96); indeed, all four lineages contain a unique mix of alleles specific to *P. barbatus* and alleles specific to *P. rugosus* at both the allozyme and microsatellite loci (Table 2), indicating hybrid ancestry. A small proportion of alleles were not found in either parent; these were virtually all (20 of 22) at highly polymorphic microsatellite loci

Table 2 Numbers of alleles specific to *P. rugosus* or *P. barbatus* found in each of the four lineages

Locus	Type	<i>P. rugosus</i>	<i>P. barbatus</i>	H1		H2		J1		J2	
				<i>P.r.</i>	<i>P.b.</i>	<i>P.r.</i>	<i>P.b.</i>	<i>P.r.</i>	<i>P.b.</i>	<i>P.r.</i>	<i>P.b.</i>
<i>PGI</i>	Protein	3	2	1	—	—	1	1	—	—	1
<i>PGM-1</i>	Protein	0	1	—	1	—	—	—	—	—	1
<i>Est-1</i>	Protein	1	2	1	1	1	1	—	—	—	—
<i>HEX</i>	Protein	1	0	1	—	1	—	—	—	—	—
<i>PGM-2</i>	Protein	0	1	—	—	—	1	—	—	—	1
<i>EST-2</i>	Protein	1	1	—	1	—	—	—	—	—	—
<i>Myrt3</i>	Microsatellite	3	1	1	—	1	—	2	—	—	1
<i>L18</i>	Microsatellite	10	12	4	—	—	3	3	2	1	5
<i>Pb5</i>	Microsatellite	7	8	—	1	—	1	3	1	1	4
<i>Pb7</i>	Microsatellite	12	3	—	2	—	—	3	—	—	2
<i>Pb8</i>	Microsatellite	3	6	1	—	—	—	1	—	1	—
Total <i>P. rugosus</i> alleles		41	—	9	—	3	—	9	—	3	—
Total <i>P. barbatus</i> alleles		—	37	—	6	7	—	—	7	—	15

Interbreeding lineages are H1 with H2 and J1 with J2. *P.r.*, *P. rugosus*; *P.b.*, *P. barbatus*.

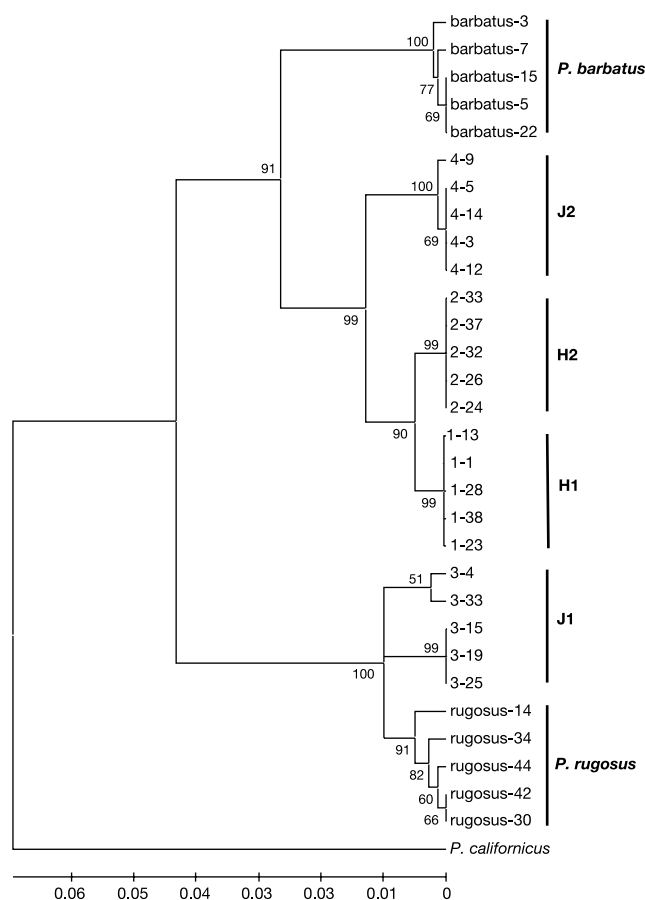


Figure 1 Linearized neighbour-joining tree of a 433-base-pair portion of the *cox1* mitochondrial gene in the two pairs of interbreeding lineages and populations of *Pogonomyrmex barbatus* and *P. rugosus*. The phylogeny generated with the Kimura two-parameter distance model (MEGA 2.1) is presented here; analyses with parsimony and maximum-likelihood methods produced the same topology. Numbers at the nodes indicate bootstrap percentages (500 replicates).

($N_a = 18\text{--}35$ alleles) and might represent unsampled parental alleles or subsequent mutations. The pattern of allelic ancestry within lineages is also consistent with hybrid origin^{11,12}. Admixture of the two species' genomes is particularly evident when comparing across loci; at any individual locus, lineages are often fixed for the allele of only one parent (Table 2 and Supplementary Information), as expected if initial polymorphisms have been lost within lineages owing to recombination and drift¹³. Thus, each of the four lineages forms a chimaeric genome, with a unique combination of the genes of the two parental species.

The chimaeric nature of the four lineages suggests a novel model for the evolution of genetic caste determination that could explain observed patterns of caste bias in both pure-lineage and interlineage offspring. In the simplest model, two loci (*A* and *B*), whose alleles differ between species (*b* is *P. barbatus* allele, *r* is *P. rugosus* allele), must interact to initiate worker development, but only conspecific alleles can interact successfully. If a hybrid lineage were fixed for alleles of different species at the two loci (either *A_bA_b/B_rB_r* or *A_rA_r/B_bB_b* genotypes), pure-lineage females would be restricted to queen development. However, correct interlocus communication would be restored in double heterozygotes (inter-lineage crosses, *A_bA_r/B_bB_r*) because each allele has a conspecific partner with which to interact. Thus, females with an inter-lineage genome would be bipotential but almost always develop into workers because too many pure-lineage females committed to queen development are

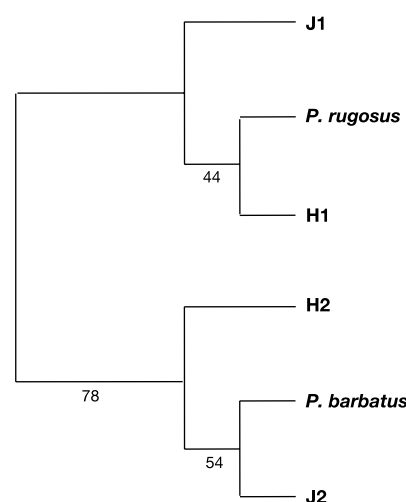


Figure 2 Maximum-likelihood consensus tree from the use of chord genetic distances²⁰ calculated from nuclear markers (six allozymes and four microsatellites). The tree is unrooted and branch lengths are not illustrated. Numbers at nodes indicate bootstrap percentages (500 replicates).

already present within colonies.

Genotypic patterns found within field colonies are consistent with the gene interaction model. Not a single pure-lineage worker has been identified in this study ($n = 75$) or in other^{6–8} studies, indicating an intrinsic breakdown in the worker caste trajectory of such offspring. Inter-lineage queens were extremely rare, both in this study ($n = 3$ of 80) and in other studies^{6,8}, and, as predicted, their occurrence seems highly dependent on the presence of pure-lineage offspring⁶. Production of inter-lineage winged queens is not distributed evenly across colonies but is restricted to certain colonies in which no pure-lineage winged queens have been observed (data for 13 colonies, 6 queens per colony; G goodness-of-fit test, $G_1 = 6.67$, $P < 0.01$). These are most probably colonies in which the queen has failed to mate with any same-lineage males⁶.

Our findings show that hybridization events between *Pogonomyrmex* species gave rise to an evolutionary novelty in which four unique hybrid lineages are linked in pairs by an unusual genetic system of caste determination. The current lack of gene flow between lineages, coupled with high levels of interbreeding, will apparently allow the system to persist indefinitely. More generally, these findings reveal a complex interaction between genetic and environmental factors in the expression of divergent caste phenotypes. □

Methods

Sampling

A single queen and worker were sampled from a total of 76 colonies from the two populations in southwestern New Mexico, USA, where the interbreeding lineages were first described: Hidalgo⁶ and Junction (~2 km south of the original *P. barbatus* site⁷). These two sites are separated by 6 km of continuous habitat. Workers were also collected from 36 colonies from a *P. barbatus* population near Blumenthal, Travis County, Texas, ~1,000 km ESE of Hidalgo and Junction, and from 44 colonies from a *P. rugosus* population in Pinal County, Arizona, ~260 km northwest of the sites (Queen Creek⁶). No heterozygosity excess, indicative of genetic caste determination, was detected in workers in either of these populations (per-locus randomization tests: *P. rugosus*, $P_{\text{LOCUS}} = 0.11\text{--}1.00$; *P. barbatus*, $P_{\text{LOCUS}} = 0.35\text{--}1.00$).

Nuclear genotyping

Allozyme protocols followed that in ref. 6. Enzyme data for 31 of the 42 Hidalgo *P. rugosus* colonies were previously presented in ref. 6; all other data are presented here. DNA was extracted from the alitrunk (= thorax) of individual ants with the Puregene tissue extraction kit (Gentra). DNA concentration was quantified in a Hoefer DyNA Quant 200 fluorimeter and diluted when necessary to achieve $10\text{--}30 \mu\text{g ml}^{-1}$ DNA. Five microsatellite loci, three designed specifically for *P. barbatus* (*Pb5*, *Pb7* and *Pb8*)¹⁴ and two designed for other genera (*Myrt3* (ref. 15) and *L18* (ref. 16)), were amplified in $10\text{-}\mu\text{l}$

reactions consisting of 1 µl DNA, 1 µl Qiagen 10 × PCR buffer, 1 µl 2 µM dNTPs, 0.8 µl 25 mM MgCl₂, 1 µl 10 µM fluorescently tagged forward and reverse primers, 0.05 µl Qiagen *Taq* polymerase and 4.15 µl sterile water. All loci were amplified with the following programme: 94 °C for 4 min, 39 cycles of 94 °C for 30 s, 54 °C or 57 °C for 30 s and 72 °C for 1.5 min, then 72 °C for 6 min and pausing at 4 °C. Loci were detected on an ABI 377 automatic sequencer.

A single winged queen was used from each colony at Hidalgo and Junction to determine allele frequencies. For most colonies (73 of 76), the multilocus genotype of the queen could be classified into one of four distinct classes or lineages (see the text). The exceptions all possessed genotypes consistent with an F₁ hybrid between two lineages (heterozygous at all diagnostic loci); such queens were noted but excluded when calculating the allele frequencies of each lineage. Workers were used to calculate allele frequencies in the *P. barbatus* and *P. rugosus* populations.

MtDNA sequencing

We analysed a 433-base-pair portion of the mitochondrial gene *cox1*, encoding cytochrome oxidase c subunit I. *Pogonomyrmex californicus* was used as the outgroup for the analysis; this species has been placed in a different complex within the genus^{17,18}. DNA was sequenced from a single winged queen or worker for five colonies of each lineage (determined from nuclear genotype) and the two parental populations. Universal insect *cox1* primers (forward, C1-J-1751; reverse, C1-N-2191) were used¹⁹, except for the substitution of A for C at position 3 of C1-N-2191. DNA was amplified in 50-µl reactions (5 µl DNA, 5 µl Qiagen 10 × buffer, 4 µl 25 mM MgCl₂, 5 µl 10 µM forward and reverse primers, 20.75 µl water, 0.25 µl Qiagen *Taq* polymerase) and amplified with the following PCR conditions: 94 °C for 4 min, 35 cycles of 94 °C for 30 s, 48 °C for 30 s and 72 °C for 1.5 min, then 72 °C for 6 min and pausing at 4 °C. PCR products were purified with Qiaquick purification columns. Sequencing reactions and detection were performed by Microsynth GmbH, Switzerland. Both forward and reverse strands were sequenced.

Received 3 February; accepted 6 May 2003; doi:10.1038/nature01744.

1. Wilson, E. O. *The Insect Societies* (Belknap, Cambridge, 1971).
2. Oster, G. F. & Wilson, E. O. *Caste and Ecology in the Social Insects* (Princeton Univ. Press, 1978).
3. Crozier, R. H. & Pamilo, P. *Evolution of Social Insect Colonies: Sex Allocation and Kin Selection* (Oxford Univ. Press, 1996).
4. Seger, J. Kinship and covariance. *J. Theor. Biol.* **91**, 191–213 (1981).
5. Wheeler, D. E. Developmental and physiological determinants of caste in social Hymenoptera—evolutionary implications. *Am. Nat.* **128**, 13–34 (1986).
6. Helms Cahan, S. et al. Extreme genetic differences between queens and workers in hybridizing *Pogonomyrmex* harvester ants. *Proc. R. Soc. Lond. B* **269**, 1871–1877 (2002).
7. Volny, V. P. & Gordon, D. M. Genetic basis for queen–worker dimorphism in a social insect. *Proc. Natl Acad. Sci. USA* **99**, 6108–6111 (2002).
8. Julian, G. E., Fewell, J. H., Gadau, J., Johnson, R. A. & Larrabee, D. Genetic determination of the queen caste in an ant hybrid zone. *Proc. Natl Acad. Sci. USA* **99**, 8157–8160 (2002).
9. Cole, A. C. *Pogonomyrmex Harvester Ants; a Study of the Genus in North America* (Univ. of Tennessee Press, Knoxville, 1968).
10. Buschinger, A. in *Social Insects—an Evolutionary Approach to Castes and Reproduction* (ed. Engels, W.) 37–57 (Springer, New York, 1990).
11. Jones, T. R., Routman, E. J., Begun, D. J. & Collins, J. P. Ancestry of an isolated subspecies of salamander, *Ambystoma tigrinum stebbinsi* Lowe: The evolutionary significance of hybridization. *Mol. Phylogenet. Evol.* **4**, 194–202 (1995).
12. Welch, M. E. & Rieseberg, L. H. Patterns of genetic variation suggest a single, ancient origin for the diploid hybrid species *Helianthus paradoxus*. *Evolution* **56**, 2126–2137 (2002).
13. Xu, S. Phylogenetic analysis under reticulate evolution. *Mol. Biol. Evol.* **17**, 897–907 (2000).
14. Volny, V. P. & Gordon, D. M. Characterization of polymorphic microsatellite loci in the red harvester ant, *Pogonomyrmex barbatus*. *Mol. Ecol. Notes* **2**, 302–303 (2002).
15. Bourke, A. F. G., Green, H. A. A. & Bruford, M. W. Parentage, reproductive skew and queen turnover in a multiple-queen ant analysed with microsatellites. *Proc. R. Soc. Lond. B* **264**, 277–283 (1997).
16. Foitzik, S., Haberl, M., Gadau, J. & Heinze, J. Mating frequency of *Leptothorax nylanderii* ant queens determined by microsatellite analysis. *Insectes Soc.* **44**, 219–227 (1997).
17. Taber, S. W. *The World of the Harvester Ants* (Texas A&M Univ. Press, College Station, 1998).
18. Parker, J. D. & Rissing, S. W. Molecular evidence for the origin of workerless social parasites in the ant genus *Pogonomyrmex*. *Evolution* **56**, 2017–2028 (2002).
19. Simon, C. et al. Evolution, weighting and phylogenetic utility of mitochondrial gene-sequences and a compilation of conserved polymerase chain-reaction primers. *Ann. Entomol. Soc. Am.* **87**, 651–701 (1994).
20. Cavalli-Sforza, L. L. & Edwards, A. W. F. Phylogenetic analysis: Models and estimation procedures. *Evolution* **21**, 550–570 (1967).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Seger for proposing the model presented here; K. Helms, G. Julian and the Student Challenge Awards Project research teams for assistance with field collections; J. Gadau and V. Volny for primer information; and M. Chapuisat, P. Christie, R. Hammond, J. Goudet and J. Parker for feedback on the manuscript. Financial support was provided by grants from the Swiss National Science Foundation to L.K. and the Durfee Foundation to S.H.C.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.H.C. (sara.helmschan@ie-zea.unil.ch).

Sustained division of the attentional spotlight

M. M. Müller*, P. Malinowski†, T. Gruber* & S. A. Hillyard‡

* Institut für Allgemeine Psychologie, Universität Leipzig, Seeburgstrasse 14–20, 04103 Leipzig, Germany

† School of Psychology, Liverpool John Moores University, 15–21 Webster Street, Liverpool L3 2ET, UK

‡ Department of Neurosciences, University of California at San Diego, 9500 Gilman Drive, La Jolla, California 92093-0608, USA

By voluntarily directing attention to a specific region of a visual scene, we can improve our perception of stimuli at that location¹. This ability to focus attention upon specific zones of the visual field has been described metaphorically as a moveable spotlight or zoom lens that facilitates the processing of stimuli within its ‘beam’^{2,3}. A long-standing controversy has centred on the question of whether the spotlight of spatial attention has a unitary beam or whether it can be divided flexibly to disparate locations^{2,4–6}. Evidence supporting the unitary spotlight view has come from numerous behavioural^{3,7–10} and electrophysiological^{11,12} studies. Recent experiments, however, indicate that the spotlight of spatial attention may be divided between non-contiguous zones of the visual field for very brief stimulus exposures (<100 ms)^{13,14}. Here we use an electrophysiological measure of attentional allocation (the steady-state visual evoked potential) to show that the spotlight may be divided between spatially separated locations (excluding interposed locations) over more extended time periods. This spotlight division appears to be accomplished at an early stage of visual-cortical processing.

To study whether the beam of spatial attention may be divided over sustained periods of several seconds, we recorded frequency-coded steady-state visual evoked potentials (SSVEPs) to concurrently presented stimuli at attended and interposed unattended locations. The SSVEP is the electrophysiological response of the visual cortex to a rapidly repeating (flickering) stimulus, and generally has a sinusoidal waveform with the same temporal frequency as the driving stimulus¹⁵. Previous studies have shown that the SSVEP amplitude is substantially increased when attention is focused upon the location of the flickering stimulus^{16,17}. The present study recorded SSVEPs to stimuli at four locations, each flickering at a different rate, so that measures of attentional allocation to spatially separated attended locations and interposed unattended locations could be obtained concurrently over periods of several seconds.

Informed consent was obtained from 15 subjects, who viewed the stimuli on a computer monitor while brain activity was recorded non-invasively from 30 scalp electrodes mounted in an elastic cap. During testing, the subject maintained fixation on a central white cross. The stimuli consisted of repetitively flashed white rectangles with superimposed red symbols that were presented at four positions along the horizontal meridian (Fig. 1). On each trial the rectangles were flashed continuously for 3.06 seconds at 15.2 Hz (position 1), 8.7 Hz (position 2), 20.3 Hz (position 3) and 12.2 Hz (position 4). Randomized sequences of five different symbols were presented at each location. Symbol presentations occurred in synchrony at the four locations with fixed durations of 181 ms.

The subject’s task was to pay attention to the symbol sequences at two of the four positions, and to push a button upon detecting the simultaneous occurrence of a particular target symbol at those two positions. On separate blocks of trials, subjects were instructed verbally to attend to either the two left field positions (1 + 2), the two right field positions (3 + 4), or to two separated positions (1 + 3) or (2 + 4). Simultaneous target symbols occurred unpre-

Table 1 Summary of behavioural results

	TDR (%)	RT (ms)	FA (%)
Attend 1 + 2	79.8 ± 3.0	472.1 ± 10.1	4.8 ± 1.0
Attend 3 + 4	77.5 ± 3.2	467.1 ± 11.7	5.3 ± 1.2
Attend 1 + 3	84.0 ± 1.7	450.7 ± 8.6	4.4 ± 0.8
Attend 2 + 4	83.8 ± 2.1	462.4 ± 8.9	5.1 ± 0.9

Target detection rates (TDR), reaction times (RT) and false alarms (FA) for the four experimental conditions. Data are shown ± standard errors.

dictably at the two attended locations (0–3 times per trial), as well as at the other locations.

The SSVEPs elicited by the flickering stimuli at the four positions were recorded concurrently, and as in previous studies^{16,18} the maximum amplitudes at each frequency were observed at occipito-temporal scalp sites contralateral to the visual field of stimulation (Fig. 1). SSVEP waveforms were generally sinusoidal, with fundamental frequencies at the driving flicker rate. The SSVEP amplitudes for each attention condition and stimulus position were quantified by complex demodulation over the three-second trial epochs^{15,16,19}. As in previous reports^{16,18} SSVEP amplitudes were generally smaller for the higher flicker rates, but for all stimuli the amplitudes were enlarged when attention was directed to their position (Fig. 2a). These amplitude increases were significant over posterior, contralateral scalp sites for position 1 ($P < 0.04$), position 2 ($P < 0.005$), position 3 ($P < 0.04$) and position 4 ($P < 0.01$), under conditions where the two positions in one visual hemifield were attended.

For investigating the division of the attentional spotlight, the critical stimulus locations were the intermediate positions 2 and 3. The SSVEP to flicker at position 2 was significantly reduced during the attend 1 + 3 condition compared to the attend 2 + 4 condition ($P < 0.007$ at electrode sites showing the largest overall attention effects); similarly, the SSVEP to flicker at position 3 was smaller during attend 2 + 4 than during attend 1 + 3 ($P < 0.0005$). These results indicate that the SSVEP to an intermediate ignored position was reduced in relation to when that same position was attended, under comparable conditions of dividing attention between separated locations. These comparisons remained significant (both $P < 0.05$) when SSVEP amplitudes were tested across three standard posterior electrode locations (PO3/4, PO7/8, O1/2).

To illustrate these effects more clearly, the SSVEP amplitudes to stimuli at positions 2 and 3 were combined following normalization with respect to each subject's maximum amplitude (Fig. 2b). The normalized amplitude at these intermediate locations was significantly reduced when attention was directed to the surrounding flanking positions (unattended intermediate) in comparison to when those same locations were attended (attended separated) ($P < 0.0001$). Importantly, the reduced amplitude at the unattended interposed positions was not significantly different from that obtained when attention was directed to the two adjacent locations in the opposite visual field ($P = \text{n.s.}$). There was also a small but significant tendency ($P < 0.05$) for the SSVEP amplitudes at attended positions 2 and 3 to be lower when attention was directed to two adjacent positions compared with two separated positions. This was paralleled in the behavioural data (Table 1) by a slightly reduced target detection rate during attention to adjacent relative to separated positions ($P < 0.02$). No significant differences were found in reaction time or false alarm rates among the different attention conditions.

To gain information about the cortical regions that mediate the divided allocation of attention, the scalp topographies of the SSVEP amplitude modulations with attention were mapped for stimulus positions 2 and 3 (Fig. 3). In both cases, the amplitude decrement for the interposed unattended location in relation to when the same stimulus was attended had a tightly focused voltage maximum over the occipital scalp contralateral to the visual field of the stimulus.

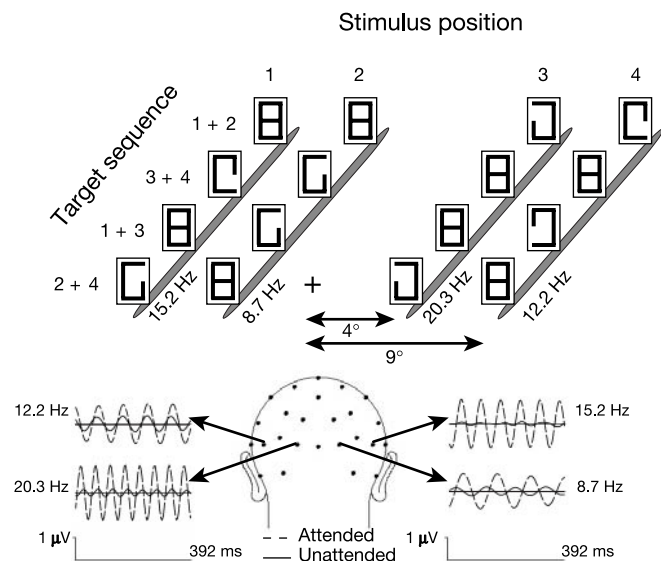


Figure 1 Schematic diagram of stimulus sequences, electrode positions and typical SSVEPs for attended (dashed lines) and unattended (solid line) conditions. Subjects reported simultaneous occurrences of the target symbol “8” at the two attended positions (either 2 + 4, 1 + 3, 3 + 4, or 1 + 2 on different blocks of trials). Stimulus sequences show examples of targets at the attended positions under the four conditions. SSVEP waveforms were obtained by a sliding average technique in the time domain^{16,25}. SSVEPs shown were obtained under conditions of attention to positions 1 + 2 and to positions 3 + 4.

These SSVEP recordings provide evidence that the spotlight of spatial attention can be divided to facilitate processing of stimuli in spatially non-contiguous locations over periods of several seconds, which extends previous findings of split attentional foci for brief exposures^{13,14}. A sustained division of attention could be inferred from the finding that SSVEP amplitudes were reduced at interposed unattended locations to the same degree as when attention was directed to adjacent stimuli in the opposite visual field. If attention had been divided only for brief, intermittent periods during task performance, one would expect less amplitude reduction at the interposed location than when attention was fully dedicated to the opposite field. The high level of accuracy at detecting the simultaneous targets in the separated positions (83–84% detection rate) with relatively few false alarms (4–5%) provides further evidence for a sustained rather than intermittent division of the attentional spotlight.

The use of brief (181 ms) target durations rules out the possibility that subjects were rapidly switching attention between the separated locations to achieve accurate performance rather than dividing attention between the locations. Although different studies have obtained varying estimates of the minimal time for switching attention, there is general agreement that the minimum time needed to identify a target at one location and then switch attention to identify a target at a second location is in the range of 200–500 ms (refs 20–24). Accordingly, the 181 ms durations used in the present study would make it impossible to achieve a high rate of target-pair detections using such a strategy.

The observed SSVEP amplitude modulations cannot be explained as a consequence of a single focus of attention being expanded to encompass the separated locations (that is, in the manner of a zoom lens). Such an expansion would necessarily include the intermediate position, and hence would result in an enhanced SSVEP for that position relative to positions outside the attended region, a pattern that was not observed. Moreover, the critical comparisons here were matched for the spatial separation of

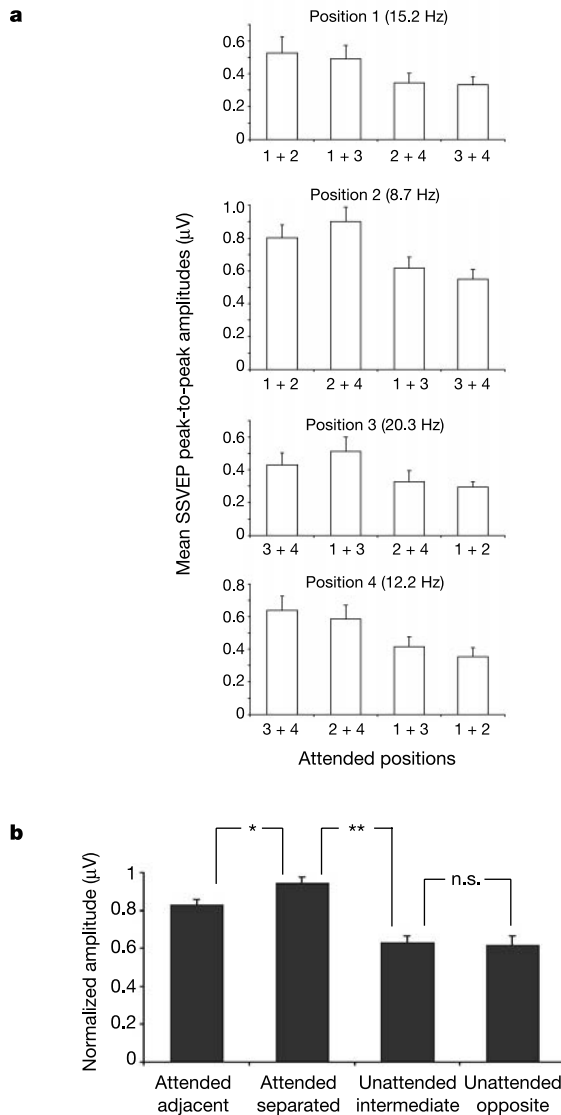


Figure 2 Mean SSVEP amplitudes at each position under the four experimental conditions and normalized amplitudes averaged across the 8.7 Hz and 20.3 Hz SSVEPs. **a**, Mean SSVEP peak-to-peak amplitudes ($\pm$ standard errors) at each position under the four experimental conditions averaged across 14 subjects. SSVEP amplitudes were obtained from electrode locations showing the largest overall attention effect. **b**, Normalized amplitude values averaged across the 8.7 Hz and 20.3 Hz SSVEPs (at positions 2 and 3, respectively) under different attention conditions. **Indicates the highly significant difference ($P < 0.0001$) between the conditions when the 8.7 Hz and 20.3 Hz stimulus sequences were attended (that is, attend 2 + 4 for 8.7 Hz and attend 1 + 3 for 20.3 Hz) as compared to conditions when these positions were intermediate and ignored (that is, attend 1 + 3 for 8.7 Hz and attend 2 + 4 for 20.3 Hz). *Indicates the significant difference ($P < 0.05$) between conditions when adjacent positions were attended (that is, attend 1 + 2 for 8.7 Hz and attend 3 + 4 for 20.3 Hz) as compared to conditions when separated positions were attended (that is, attend 2 + 4 for 8.7 Hz and attend 1 + 3 for 20.3 Hz). No significant differences were found when comparing unattended intermediate positions versus unattended positions in the hemifield opposite to the attended adjacent positions.

the attended locations, and both the behavioural and SSVEP results suggest that there was no cost for attending to two separated positions in comparison with attending to two adjacent positions (indeed, a slight advantage for the separated positions was obtained). This contrasts with findings of reduced performance during attention to larger areas as predicted by the zoom lens model³, and raises questions about the applicability of this model to

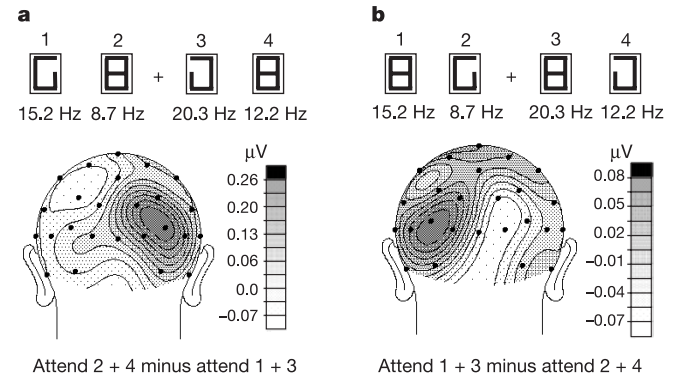


Figure 3 Spline-interpolated isocontour maps of the grand average attend minus unattend amplitude under conditions of attending to separated locations. Data are shown for the 8.7 Hz SSVEP amplitude at position 2 (**a**) and the 20.3 Hz SSVEP at position 3 (**b**). Note the different voltage scales for the two maps.

continuously presented stimuli as in the present design. Previous studies of SSVEPs have similarly found evidence for highly flexible allocations of attention to flickering stimulus patterns (for example, to stimuli in the form of a ring, excluding the centre²⁵). This suggests that spatial allocation of attention may be governed by different principles for continuously presented (for example, flickering) stimuli than for briefly flashed stimuli, and it seems reasonable to propose that the former may more closely simulate the natural environment of relatively permanent objects.

The lateral occipital scalp topography of the SSVEP modulations during divided attention is consistent with neural activity localized to extrastriate visual cortical areas of the occipital lobe^{16,26,27}. Previous studies using both steady-state and transient stimulation have found that stimuli falling within the spotlight of attention elicit enhanced neural responses in extrastriate visual cortex^{16,28,29}, and that this facilitation of attended signals occurs early in visual processing, with a latency or phase lag of 80–140 ms following stimulus onsets^{17,29}. The present findings provide evidence that the spotlight may be divided for sustained periods, thereby facilitating inputs from separated locations at the level of the extrastriate occipital pathways. This would appear to be a highly efficient early selection mechanism for distributing attention optimally to dispersed stimuli in the visual surroundings. □

Methods

Stimuli

Stimuli were presented against a dark background at four different locations, lined up along the horizontal meridian of a 19-inch computer monitor set to a resolution of 800×600 pixels with a refresh rate of $60.8 \text{ frames s}^{-1}$. At a viewing distance of 70 cm, each individual white rectangle subtended a viewing angle of $2.5 \times 3.2^\circ$ with an eccentricity of 4° for the medial and 9° for the lateral rectangles, respectively (distance between the inner edges of the elements to the central fixation cross). As depicted in Fig. 1, the rectangles at positions 1 to 4 flickered at rates of 15.20 Hz, 8.69 Hz, 20.27 Hz and 12.16 Hz, respectively, corresponding to cycle durations of 4, 7, 3 and 5 frames. Each rectangle was 'on' for 1 frame (16.45 ms), followed by the appropriate number of 'off' frames to achieve the respective stimulation frequency (for example, one 'on' frame followed by 3 'off' frames for 15.20 Hz). Five different symbols were used, with one symbol serving as the target. The symbols were presented simultaneously at the 4 locations for 11 frames (180.95 ms) followed immediately by the next symbol array. Thus, the symbol onsets and offsets did not occur in synchrony with the background flickering rectangles that drove the SSVEP. Target symbols occurred equally often at all stimulus positions, following randomized sequences.

Experimental procedure

Each trial lasted 3,060 ms, and perfect synchronization of all four flicker frequencies was set at 526 ms after flicker-onset, which served as reference point for extracting epochs for electroencephalogram (EEG) analysis. The fixation cross and four rectangles that outlined the stimulus positions remained visible during the inter-trial intervals. On about 75% of the trials, between one and three target pairs (simultaneous occurrences at the two attended locations) were randomly presented. Paired targets were separated by a

minimum interval of 905 ms (55 frames). The experiment consisted of 12 blocks of 40 trials each, resulting in 120 trials per experimental condition. Blocks were administered in random order. The responding hand was changed halfway through the experiment, and the sequence of hand usage was counterbalanced across subjects. One subject was excluded from further analysis because of difficulty in detecting the matching symbols.

Data acquisition

The electrode positions shown schematically in Fig. 1 are specified in a previous report³⁰. The EEG was recorded with a sampling rate of 250 Hz and a bandpass of d.c. to 50 Hz. Individual trials were discarded on the basis of blink or electromyogram (EMG) artefacts in the scalp channels exceeding 75 μ V, or when lateral eye movements monitored in the horizontal electro-oculogram (EOG) deviated more than 11 μ V (1°) from fixation. These stringent criteria resulted in a mean rejection rate of 30% of the trials. In order to analyse the SSVEP, artefact-free EEG epochs were averaged separately for each experimental condition and algebraically re-referenced to averaged mastoids by subtracting one-half of the averaged signal recorded from the mastoid opposite the reference mastoid from the averaged signals at each scalp site. The averaging epochs extended from 500 ms before to 2,500 ms after the time point when all streams were synchronized (that is, from 26 to 3,026 ms after flicker onset). SSVEP amplitudes were extracted by means of complex demodulation of the averaged waveforms^{15,19,25}. To avoid including the visual evoked response to flicker onset in the SSVEP measurements, the first 500 ms of each epoch were excluded from analysis. Thus, mean SSVEP amplitudes were calculated over the interval 526 to 3,026 ms after flicker onset.

Data analysis

For testing the significance of SSVEP amplitude changes, the posterior electrode site that exhibited the largest overall attention effect (comparing attended versus ignored positions) was selected for each subject. These amplitude values were subjected to paired *t*-tests between conditions, and corrected for multiple comparisons by the Bonferroni–Dunn criterion. As a reliability check, the averaged amplitudes across three standard electrode locations (PO3/4, PO7/8 and O1/2) were also subjected to paired *t*-tests for the attended versus ignored SSVEPs at 8.69 Hz and 20.27 Hz under conditions of attention to separated locations.

Only button-presses occurring between 250 and 1,000 ms after target-pair onset were accepted as correct detections. False alarms for the adjacent condition were defined as button presses occurring in response to a target presented at only one of the attended locations. For the separate conditions, false alarms were defined as button presses in response to a target presented in only one of the attended locations and/or in the intermediate to-be-ignored position. Target detection rates, reaction times and false alarms were tested by one factor repeated measures analysis of variance (experimental condition).

Received 3 April; accepted 27 May 2003; doi:10.1038/nature01812.

1. LaBerge, D. *Attentional Processing* (Harvard Univ. Press, Cambridge, Massachusetts, 1995).
2. Posner, I. P. & Petersen, S. E. The attention system of the human brain. *Annu. Rev. Neurosci.* **13**, 25–42 (1990).
3. Eriksen, C. W. & Yeh, Y. Y. Allocation of attention in the visual field. *J. Exp. Psychol. Hum. Percept. Perform.* **11**, 583–597 (1985).
4. LaBerge, D. & Brown, V. Theory of attentional operations in shape identification. *Psychol. Rev.* **96**, 101–124 (1989).
5. Shaw, M. L. & Shaw, P. Optimal allocation of cognitive resources to spatial locations. *J. Exp. Psychol. Hum. Percept. Perform.* **3**, 201–211 (1977).
6. Castiello, U. & Umiltà, C. Splitting focal attention. *J. Exp. Psychol. Hum. Percept. Perform.* **18**, 837–848 (1992).
7. Posner, M. I., Snyder, C. R. R. & Davidson, B. J. Attention and detection of signals. *J. Exp. Psychol. Gen.* **109**, 160–174 (1980).
8. Pan, K. & Eriksen, C. W. Attentional distribution in the visual field during same-different judgements as assessed by response competition. *Percept. Psychophys.* **53**, 134–144 (1993).
9. McCormick, P. A., Klein, R. M. & Johnston, S. Splitting vs. shared visual attention: An empirical commentary on Castiello & Umiltà (1992). *J. Exp. Psychol. Hum. Percept. Perform.* **24**, 350–357 (1998).
10. Kiefer, R. J. & Siple, P. Spatial constraints on the voluntary control of attention across visual space. *Can. J. Psychol.* **41**, 474–489 (1987).
11. Eimer, M. Attending to quadrants and ring-shaped regions: ERP effects of visual attention in different spatial selection tasks. *Psychophysiology* **36**, 491–503 (1999).
12. Heinze, H.-J. et al. Attention to adjacent and separate positions in space: An electrophysiological analysis. *Percept. Psychophys.* **56**, 42–52 (1994).
13. Awh, E. & Pashler, H. Evidence for split attentional foci. *J. Exp. Psychol. Hum. Percept. Perform.* **26**, 834–846 (2000).
14. Hahn, S. & Kramer, A. E. Further evidence for the division of attention between noncontiguous locations. *Vis. Cognit.* **5**, 217–256 (1998).
15. Regan, D. *Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine* (Elsevier, New York, 1989).
16. Müller, M. M. et al. Effects of spatial selective attention on the steady-state visual evoked potential in the 20–28 Hz range. *Cognit. Brain Res.* **6**, 249–261 (1998).
17. Di Russo, F. & Spinelli, D. Spatial attention has different effects on the magno- and parvocellular pathways. *NeuroReport* **10**, 2755–2762 (1999).
18. Müller, M. M., Teder-Sälejärvi, W. & Hillyard, S. A. The time course of cortical facilitation during cued shifts of spatial attention. *Nature Neurosci.* **1**, 631–634 (1998).
19. Müller, M. M. et al. in *Oscillatory Event-related Brain Dynamics* (eds Pantev, C., Elbert, T. & Lütkenhöner, B.) 325–342 (Plenum, New York, 1994).
20. Reeves, A. & Sperling, G. Attention gating in short-term visual memory. *Psychol. Rev.* **93**, 180–206 (1986).

21. Weichselgartner, E. & Sperling, G. Dynamics of automatic controlled visual attention. *Science* **238**, 778–780 (1987).
22. Duncan, J., Ward, R. & Shapiro, K. Direct measurement of attentional dwell time in human vision. *Nature* **369**, 313–315 (1994).
23. Peterson, M. S. & Juola, J. F. Evidence for distinct attentional bottlenecks in attention switching and attentional blink tasks. *J. Gen. Psychol.* **127**, 6–26 (2000).
24. Moore, C. M., Egeth, H., Berglan, L. & Luck, S. J. Are attentional dwell times inconsistent with serial visual search? *Psychonom. Bull. Rev.* **3**, 360–365 (1996).
25. Müller, M. M. & Hübner, R. Can the attentional spotlight be shaped like a doughnut? Evidence from steady state visual evoked potentials. *Psychol. Sci.* **13**, 119–124 (2002).
26. Müller, M. M., Teder, W. & Hillyard, S. A. Magnetoencephalographic recording of steady-state visual evoked cortical activity. *Brain Topogr.* **9**, 163–168 (1997).
27. Hillyard, S. A. et al. Combining steady-state visual evoked potentials and fMRI to localize brain activity during selective attention. *Hum. Brain Mapp.* **5**, 287–292 (1997).
28. Clark, V. P. & Hillyard, S. A. Spatial selective attention affects early extrastriate but not striate components of the visual evoked potential. *J. Cognit. Neurosci.* **8**, 387–402 (1996).
29. Martinez, A. et al. Putting spatial attention on the map: Timing and localization of stimulus selection processes in striate and extrastriate visual areas. *Vis. Res.* **41**, 1437–1457 (2001).
30. Clark, V. P., Fan, S. & Hillyard, S. A. Identification of early visual evoked potential generators by retinotopic and topographic analyses. *Hum. Brain Mapp.* **2**, 170–187 (1995).

Acknowledgements We thank N. Williams, H. Messmer and C.-M. Giabbiconi for help in data recording. This work was supported by the Deutsche Forschungsgemeinschaft and by NIMH.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.M.M. (m.mueller@rz.uni-leipzig.de).

Perceptual consequences of centre-surround antagonism in visual motion processing

Duje Tadin, Joseph S. Lappin, Lee A. Gilroy & Randolph Blake

Vanderbilt Vision Research Center, Vanderbilt University, 111 21st Avenue South, Nashville, Tennessee 37203, USA

Centre-surround receptive field organization is a ubiquitous property in mammalian visual systems, presumably tailored for extracting image features that are differentially distributed over space¹. In visual motion, this is evident as antagonistic interactions between centre and surround regions of the receptive fields of many direction-selective neurons in visual cortex^{2–6}. In a series of psychophysical experiments we make the counterintuitive observation that increasing the size of a high-contrast moving pattern renders its direction of motion more difficult to perceive and reduces its effectiveness as an adaptation stimulus. We propose that this is a perceptual correlate of centre-surround antagonism, possibly within a population of neurons in the middle temporal visual area. The spatial antagonism of motion signals observed at high contrast gives way to spatial summation as contrast decreases. Evidently, integration of motion signals over space depends crucially on the visibility of those signals, thereby allowing the visual system to register motion information efficiently and adaptively.

Centre-surround neurons, especially those in the middle temporal visual area (MT), are believed to be crucially involved in the perception of object motion^{3,7}, in figure-ground segmentation^{7–9} and in the registration of three-dimensional shape from motion^{9,10}. By analogy with other aspects of vision¹¹, if centre-surround antagonism is an integral part of motion processing, we should expect to see a perceptual signature of this antagonism in the form of impaired motion visibility with increasing stimulus size. However, existing evidence shows that increasing the size of a low-contrast moving stimulus enhances its visibility^{12,13}, presumably

owing to spatial summation. Such psychophysical estimations of the spatial properties of motion mechanisms tend to be based on low-contrast or noisy stimuli, whereas physiological descriptions of centre-surround motion neurons have been obtained with high-contrast motion. Moreover, in the visual cortex, the nature of centre-surround interactions is often dependent on contrast: surround suppression is stronger at high contrast and spatial summation is more pronounced at low contrast^{14–16}. Thus, threshold contrast measurements might not fully describe the spatial properties of human motion perception, especially at high contrast. With this in mind, we devised alternative approaches for measuring motion sensitivity.

In our first experiment we measured the threshold exposure duration required for human observers to accurately identify the motion direction of a drifting Gabor patch. In separate conditions, observers viewed foveally presented Gabor patches of various sizes and contrasts. Spatial frequency and speed were fixed at 1 cycle per degree and 2° s^{-1} , respectively. The contrast of a Gabor patch was ramped on and off with a temporal gaussian envelope, allowing the presentation of brief motion stimuli.

Increasing stimulus contrast resulted in a marked change in the way in which motion signals were integrated over space (Fig. 1). At low contrast (2.8%), duration thresholds decreased with increasing size, reaching a lower asymptote at about 40 ms (Fig. 1a, c). This

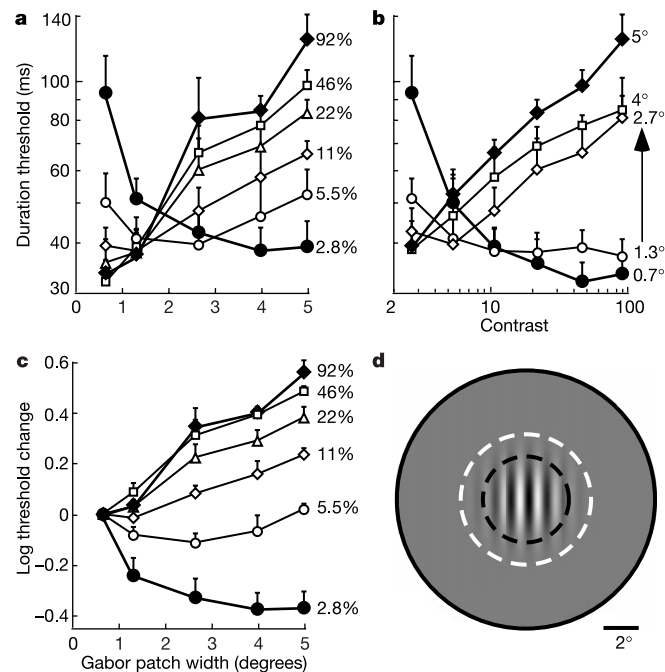


Figure 1 Effects of size and contrast on motion perception. Individual data points are averages for five observers. Results are means $\pm$ s.e.m. **a**, Duration thresholds as a function of stimulus size at different contrasts. **b**, Duration thresholds as a function of contrast for a range of stimulus sizes. **c**, Log of threshold change as a function of stimulus size at different contrasts. For each observer, the logarithm of the threshold change was calculated relative to the duration threshold for the smallest size (0.7°) at each contrast level. Note that the transition from suppression to summation occurs at about 5% contrast, a value that, coincidentally or not, is the contrast in which MT neurons attain about 25% of their maximum response on average²⁸. **d**, A Gabor patch 2.7° wide shown relative to an average foveal macaque MT receptive field. The dashed dark circle illustrates the stimulus size beyond which an average foveal MT centre-surround neuron exhibits surround suppression⁶. The radius of the surround is usually about three times the centre radius, as indicated by the full circle. The full spatial extent of the Gabor patch ($r = 3\sigma = 4^\circ$) is indicated by the white dashed circle. This comparison assumes that the properties of human and macaque MT are comparable²⁹ and that the receptive field sizes are similar for the two species³⁰.

result, implying spatial summation of motion signals, is consistent with earlier reports^{12,13}. At high contrast, however, duration thresholds increased fourfold as the Gabor patch width increased from 0.7° to 5° . In other words, for small Gabor patches, performance improved with increasing contrast, whereas for large Gabor patches, performance deteriorated substantially with contrast (Fig. 1b). These highly surprising results imply neural processes fundamentally different from spatial summation. Closer examination of the results reveals that the increase in duration threshold was greatest for Gabor patches larger than 2.7° in width (arrow in Fig. 1b), indicating the existence of a 'critical size'. We speculated that surround inhibition might be responsible for the observed decrease in motion sensitivity, leading us to explore this hypothesis in several more experiments.

Similar contrast-dependent size effects were obtained with faster-moving stimuli (8° s^{-1}) and with Gabor patches whose spatial bandwidth was held constant by scaling spatial frequency (1 cycle per σ). We also manipulated the effective stimulus contrast by adding variable amounts of dynamic noise to a fixed-contrast Gabor patch. We found evidence for spatial summation when motion appeared within high-contrast noise, and evidence for spatial suppression when motion appeared within weak noise or when noise was absent altogether (Fig. 2a). In other words, the presence of noise actually improved the visibility of large-motion stimuli.

Most neurophysiological explorations of centre-surround motion neurons have been performed with spatially broadband random-dot displays. We therefore also investigated effects of size with random-dot stimuli (light and dark pixels, each 3×3 arcmin) presented in a spatial gaussian envelope. From frame to frame of the animation, half of the pixels shifted in one direction (6.2° s^{-1}) while the remaining pixels were randomly regenerated, conditions producing vivid motion perception at suprathreshold exposure durations. Duration thresholds with these stimuli, too, yielded evidence for spatial summation at low contrast and spatial suppression at high contrast (Fig. 2b).

Realizing that the receptive field sizes of motion-sensitive neurons increase with retinal eccentricity^{6,17}, we wondered whether the detrimental effect of stimulus size at high contrast would diminish with increasing eccentricity. Accordingly, we manipulated the display size for a range of eccentricities with contrast fixed at 92%. Once again, foveal presentation yielded evidence for surround suppression (Fig. 2c). As eccentricity increased, duration thresholds decreased for all sizes. More importantly, the size dependence of duration thresholds changed systematically with eccentricity, with almost no effect at the largest eccentricity tested.

The motion strength of a periodic Gabor patch can also be varied

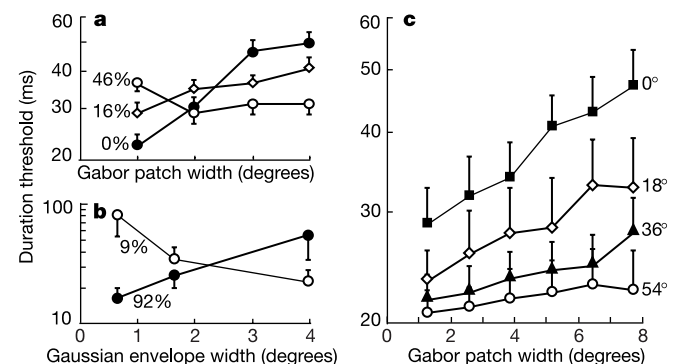


Figure 2 Results from added-noise, random-dot and eccentricity experiments. **a**, Duration thresholds as a function of stimulus size at different noise contrasts. **b**, Duration thresholds as a function of random-dot stimulus size at different contrasts. **c**, Duration thresholds as a function of stimulus size at different eccentricities. Results are means $\pm$ s.e.m.

by adjusting the magnitude of an abrupt phase shift¹⁸: increasing phase shifts from 0° to 90° enhances the visibility of motion. We conducted an experiment in which observers identified the motion direction of a fixed-contrast Gabor patch that abruptly shifted in phase in the middle of a 100-ms presentation interval. Threshold phase shift was obtained for a range of contrasts and sizes with spatial frequency fixed at 0.5 cycle per degree. Results (Fig. 3a) replicated the duration threshold findings, again showing the spatial suppression of motion signals at high contrast.

Given the reduced visibility of isoluminant moving stimuli¹⁹, we might expect spatial summation of isoluminant stimuli similar to that observed for low-contrast and noisy stimuli. Using the phase-shift procedure from the previous experiment, we compared thresholds for isoluminant (red–green) and high-luminance contrast (yellow–black) Gabor patches. Once again, luminance contrast stimuli exhibited surround suppression (Fig. 3b). However, thresholds for isoluminant stimuli decreased with increasing size, exhibiting spatial summation. This result is surprising in that colour-defined motion with large stimuli was actually perceived more accurately than luminance-defined motion, presumably because the latter is affected by surround suppression.

Surround inhibition in motion-sensitive neurons decreases neural activity in response to a large moving stimulus. This decreased activity might be evident in weaker adaptation of motion-sensitive centre-surround neurons during prolonged exposure to motion. The motion after-effect (MAE)—an illusory perception of motion after prolonged exposure to motion—is thought to reflect the adaptation of motion-sensitive neurons²⁰. Given our findings, we predicted that adapting to a large high-contrast moving stimulus should result in a weaker MAE, whereas adaptation to a low-contrast stimulus should not. We explored these predictions by inducing MAE with moving Gabor patches of various sizes and contrasts, and testing MAE strength with a small test stimulus. As predicted, MAE strength decreased with increasing size when observers adapted to a high-contrast Gabor patch, indicating spatial suppression (Fig. 4). This result is consistent with earlier observations^{21,22} that were restricted to high-contrast adapting stimuli. In contrast, as predicted, when a low-contrast adapting stimulus was used, MAE strength increased with increasing size.

The present study reveals that large-sized objects detrimentally affect human motion perception, which contradicts intuition and challenges accepted ideas about the spatial properties of motion perception. In psychophysics, spatial summation has often been assumed as a basic characteristic of motion processing^{12,13}, but we show here that this holds only for conditions of low visibility. In addition, our results corroborate and help to explain some earlier

findings. Dividing a large high-contrast object into smaller parts actually improved performance in a speed discrimination task²³, leading to the suggestion of surround suppression as one possible explanation. Increasing the contrast of a briefly presented, large drifting grating decreased performance²⁴, agreeing with a portion of our findings (filled diamonds in Fig. 1b). Brief motion stimuli, by virtue of their broad temporal frequency spectrum, might therefore stimulate motion filters tuned to opposing directions of motion²⁴. At high contrast, these paired filters could saturate impairing direction identification; however, this fails to explain the effects of size that are central to the present study. Moreover, paired opponent filters cannot explain the MAE results (Fig. 4) that generalize our principal findings to prolonged motion stimuli.

For several reasons we believe that our results might reflect the receptive field properties of centre-surround neurons in MT. First, impaired visual performance with larger stimuli has been construed as the perceptual signature expected from antagonistic centre-surround mechanisms¹¹. Second, the ‘critical size’ at which we begin to observe strong surround suppression (Fig. 1b) is large enough to impinge on the surrounds of MT neurons with foveal receptive fields (Fig. 1d). However, this critical size is much larger than receptive fields in the primary visual area (V1) and much smaller than receptive fields in the lateral part of area MST (medial superior temporal), which are cortical areas other than MT that contain centre-surround motion neurons^{4,5}. Third, the detrimental effect of stimulus size diminishes in the visual periphery, which is consistent with the increase in sizes of MT receptive fields with eccentricity^{6,17}. Fourth, MAE, a perceptual after-effect attributed, at least in part, to MT activity²⁰, is weaker if induced with large high-contrast stimuli. This result is expected if such stimuli inhibit the activity of MT neurons whose adaptation normally contributes to the MAE. Last, MT neurons respond more weakly to motion of isoluminant gratings than to motion of luminance gratings²⁵, a property that dovetails nicely with the failure of isoluminant motion to produce surround suppression.

Our conclusion rests on several assumptions, which are not unreasonable given existing evidence. We assume that the quality of motion perception covaries with underlying neuronal firing rate—a reasonable assumption for MT neurons²⁶. We also assume that the strength of surround suppression induced by a large moving object is not substantially altered because of the variations in receptive field size and eccentricity. Of course, surrounds of some neurons will be only partly stimulated, particularly those with receptive field centres aligned along the stimulus border. However, because the border of our stimuli is blurred and low in contrast (Fig. 1d), these ‘border neurons’, too, will be affected by surround

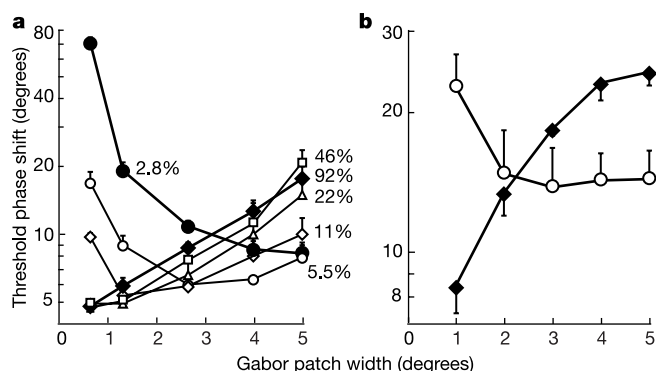


Figure 3 Results from phase-shift and isoluminant-motion experiments. **a**, Phase-shift thresholds as a function of stimulus size at different contrasts. **b**, Phase-shift thresholds as a function of stimulus size for luminance contrast (filled circles) and isoluminant (open circles) stimuli. Results are means $\pm$ s.e.m.

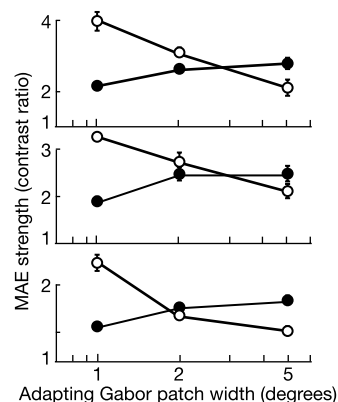


Figure 4 Effects of stimulus size and contrast on the MAE strength for three subjects, A.P. (top), R.B. (middle) and D.T. (bottom). Filled symbols show MAE strength for low-contrast adaptation (2.8%). Empty symbols show results for high-contrast adaptation (26%). Results are means $\pm$ s.e.m.

suppression, because their partly stimulated surrounds will receive higher contrast stimulation than their centres. Last, we assume that our results arise from centre-surround neurons, and not the so-called 'wide-field' MT neurons, which show no surround suppression³. Centre-surround and wide-field neurons are anatomically segregated^{3,6} and produce different behavioural effects when artificially stimulated, leading to the conclusion that wide-field neurons might not be directly involved in signalling object motion⁷.

Our results generate testable predictions. First, they predict that response strength in MT to isoluminant motion stimuli, although weak, should increase with stimulus size. Second, the observed contrast dependence of the spatial integration of motion signals is particularly interesting. It predicts contrast-dependent changes in receptive field properties of MT neurons similar to those observed in V1 (refs 14–16), which should be measurable with neurophysiological and neuroimaging techniques.

Why, though, is the nature of centre-surround interactions in motion perception so markedly affected by contrast? At high contrast, the computational benefits of surround suppression^{7–10} probably outweigh the obligatory decrease in neuronal activity and reduced sensitivity. At low contrast, high sensitivity is essential, so it makes functional sense that receptive field organization shifts from surround suppression to spatial summation. The integration of motion signals over space is therefore an adaptive process that enables the visual system to process moving stimuli more efficiently by employing computationally important suppressive mechanisms only when the sensory input is strong enough to guarantee visibility. □

Methods

Displays and viewing conditions

Patterns were shown on a monitor with linear display characteristics (800 pixels × 600 pixels resolution, 120 Hz). The main results were confirmed at 160 and 200 Hz. Viewing was binocular at 83 cm. The ambient illumination was 4.8 cd m⁻². The background luminance was 60.5 cd m⁻². The size was defined as the 2σ width of a Gabor patch, a drifting vertical sine grating windowed by a stationary two-dimensional gaussian envelope (Fig. 1d), where σ is the standard deviation of the gaussian. Duration was defined as 2σ of the temporal gaussian envelope. On each trial, a drifting Gabor patch was presented foveally and observers indicated the perceived direction (left or right) by a key press. Feedback was provided. Duration thresholds (82%) were estimated by interleaved Quest staircases. For each condition, five observers ran four pairs of interleaved staircases; the first pair was discarded as practice. All experiments complied with institutionally reviewed procedures for human subjects.

Visual noise experiment

Speed was increased to 4° s⁻¹. A 20° by 20° noise pattern consisted of 480 × 480 dark and light pixels and was randomly regenerated at 120 Hz. Noise contrast was defined as the luminance difference between light and dark pixels divided by the sum of their luminances. This noise mask was combined additively with a foveally presented Gabor patch (46% contrast). Other methods were the same as in the first experiment.

Eccentric presentation experiment

To present visually larger stimuli within the limited monitor area, a gaussian spatial envelope was replaced with the two-dimensional raised cosine envelope. Size was defined as the distance between two diametrically opposing points on the raised cosine envelope where the contrast was 60.7% of the peak contrast (analogous to the 2σ of a gaussian distribution). Spatial frequency was lowered to 0.5 cycle per degree and the speed was increased to 4° s⁻¹. In this experiment only, motion directions were vertical (up or down). Other methods were the same as in the first experiment.

Isoluminant motion experiment

The Gabor patch consisted of spatially overlapping isoluminant red and green gratings. For each observer, the red-green isoluminant point was obtained with the minimum-motion technique²⁷. If isoluminant red and green gratings are presented spatially in antiphase, the resulting red-green compound grating has no luminance contrast. Presenting the same two gratings in phase produces a yellow-black luminance grating. The spatial frequency was 1 cycle per degree. Results similar to those in Fig. 3b were obtained by measuring duration thresholds.

MAE experiment

Observers ($n = 3$) adapted to a foveally centred Gabor patch (1 cycle per degree) moving at 4° s⁻¹. The test stimulus consisted of two overlapping Gabor patches drifting in opposite directions (the stimulus parameters were the same as in the adapting Gabor patch except the width, which was fixed to 1°). When the contrasts of two Gabor patches were identical, the motion was ambiguous and the test stimulus seemed to flicker. Adapting to a

motion in one direction effectively decreases the motion strength of the Gabor patch moving in the adapted direction and the test stimulus appears to move in the opposite direction. Then, to restore the perception of flicker, the contrast of the Gabor patch moving in the adapted direction is increased and the contrast of the other Gabor patch is decreased. The contrast ratio required to restore flicker perception is taken as a measure of MAE strength (higher contrast ratio corresponds to stronger MAE). The initial adaptation was for 30 s (10 s 'top-off' adaptation after the first trial), followed by a 0.3-s blank screen and a 1-s test stimulus. After viewing the test stimulus, observers indicated the perceived direction. The contrast ratio of two Gabor patches composing the test stimulus was then adjusted under the control of two interleaved 'one-up—one-down' staircases. Each staircase converged after six reversals, with the average of the last four reversals taken as the result. Eight measurements were made for each condition.

Received 15 January; accepted 9 May 2003; doi:10.1038/nature01800.

1. Allman, J., Miezin, F. & McGuinness, E. Stimulus specific responses from beyond the classical receptive field: Neurophysiological mechanisms for local-global comparisons in visual neurons. *Annu. Rev. Neurosci.* **8**, 407–430 (1985).
2. Allman, J., Miezin, F. & McGuinness, E. Direction- and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area (MT). *Perception* **14**, 105–126 (1985).
3. Born, R. T. & Tootell, R. B. Segregation of global and local motion processing in primate middle temporal visual area. *Nature* **357**, 497–499 (1992).
4. Eifuku, S. & Wurtz, R. H. Response to motion in extrastriate area MSTl: Center-surround interactions. *J. Neurophysiol.* **80**, 282–296 (1998).
5. Jones, H. E., Grieve, K. L., Wang, W. & Sillito, A. M. Surround suppression in primate V1. *J. Neurophysiol.* **86**, 2011–2028 (2001).
6. Raiguel, S. E., van Hulle, M. M., Xiao, D. K., Marcar, V. L. & Orban, G. A. Shape and spatial distribution of receptive fields and antagonistic motion surround in the middle temporal area (V5) of the macaque. *Eur. J. Neurosci.* **7**, 2064–2082 (1995).
7. Born, R. T., Groh, J. M., Zhao, R. & Lukasewycz, S. J. Segregation of object and background motion in visual area MT: Effects of microstimulation on eye movements. *Neuron* **26**, 725–734 (2000).
8. Nakayama, K. & Loomis, J. M. Optical velocity patterns, velocity-sensitive neurons, and space perception: A hypothesis. *Perception* **3**, 63–80 (1974).
9. Gautama, T. & Van Hulle, M. M. Function of center-surround antagonism for motion in visual area MT/V5: A modeling study. *Vision Res.* **41**, 3917–3930 (2001).
10. Buracas, G. T. & Albright, T. D. Contribution of area MT to perception of three-dimensional shape: Computational study. *Vision Res.* **36**, 869–887 (1996).
11. Westheimer, G. Spatial interaction in human cone vision. *J. Physiol. (Lond.)* **190**, 139–154 (1967).
12. Anderson, S. J. & Burr, D. C. Spatial summation properties of directionally sensitive mechanisms in human vision. *J. Opt. Soc. Am. A* **8**, 1330–1339 (1991).
13. Watson, A. B. & Turano, K. The optimal motion stimulus. *Vision Res.* **35**, 325–336 (1995).
14. Kapadia, M. K., Westheimer, G. & Gilbert, C. D. Dynamics of spatial summation in primary visual cortex of alert monkeys. *Proc. Natl Acad. Sci. USA* **96**, 12073–12078 (1999).
15. Levitt, J. B. & Lund, J. S. Contrast dependence of contextual effects in primate visual cortex. *Nature* **387**, 73–76 (1997).
16. Sceniak, M. P., Ringach, D. L., Hawken, M. J. & Shapley, R. Contrast's effect on spatial summation by macaque V1 neurons. *Nature Neurosci.* **2**, 733–739 (1999).
17. Albright, T. D. Direction and orientation selectivity of neurons in visual area MT of the macaque. *J. Neurophysiol.* **52**, 1106–1130 (1984).
18. Nakayama, K. & Silverman, G. H. Detection and discrimination of sinusoidal grating displacements. *J. Opt. Soc. Am. A* **2**, 267–274 (1985).
19. Dobkins, K. R. & Albright, T. D. In *High-level Motion Processing* (ed. Watanabe, T.) 53–94 (MIT Press, Cambridge, Massachusetts, 1998).
20. Huk, A. C., Ress, D. & Heeger, D. J. Neuronal basis of the motion aftereffect reconsidered. *Neuron* **32**, 161–172 (2001).
21. Murakami, I. & Shimojo, S. Modulation of motion aftereffect by surround motion and its dependence on stimulus size and eccentricity. *Vision Res.* **35**, 1835–1844 (1995).
22. Sachtler, W. L. & Zaidi, Q. Effect of spatial configuration on motion aftereffects. *J. Opt. Soc. Am. A* **10**, 1433–1449 (1993).
23. Verghese, P. & Stone, L. S. Perceived visual speed constrained by image segmentation. *Nature* **381**, 161–163 (1996).
24. Derrington, A. M. & Goddard, P. A. Failure of motion discrimination at high contrasts: Evidence for saturation. *Vision Res.* **29**, 1767–1776 (1989).
25. Gegenfurtner, K. R. *et al.* Chromatic properties of neurons in macaque MT. *Vis. Neurosci.* **11**, 455–466 (1994).
26. Britten, K. H., Shadlen, M. N., Newsome, W. T. & Movshon, J. A. The analysis of visual motion: A comparison of neuronal and psychophysical performance. *J. Neurosci.* **12**, 4745–4765 (1992).
27. Cavanagh, P., MacLeod, D. I. & Anstis, S. M. Equiluminance: Spatial and temporal factors and the contribution of blue-sensitive cones. *J. Opt. Soc. Am. A* **4**, 1428–1438 (1987).
28. Sclar, G., Maunsell, J. H. & Lennie, P. Coding of image contrast in central visual pathways of the macaque monkey. *Vision Res.* **30**, 1–10 (1990).
29. Rees, G., Friston, K. & Koch, C. A direct quantitative relationship between the functional properties of human and macaque V5. *Nature Neurosci.* **3**, 716–723 (2000).
30. Kastner, S. *et al.* Modulation of sensory suppression: Implications for receptive field sizes in the human visual cortex. *J. Neurophysiol.* **86**, 1398–1411 (2001).

Acknowledgements We thank J. Schall, B. Borghuis, S. Shorter-Jacobi and A. Panduranga for comments on the experiments and manuscript. This work was supported by a grant from the NIH.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.T. (duj.tadin@vanderbilt.edu).

Analgesia and hyperalgesia from GABA-mediated modulation of the cerebral cortex

Luc Jasmin^{*†}, Samuel D. Rabkin[‡], Alberto Granato[§], Abdennacer Boudah^{*} & Peter T. Ohara[†]

^{*} Department of Neurological Surgery and [†] Department of Anatomy, University of California, San Francisco, San Francisco, California 94143-0452, USA

[‡] Department of Neurosurgery, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02129, USA

[§] Department of Psychology, Catholic University, 20123 Milan, Italy

It is known that pain perception can be altered by mood, attention and cognition, or by direct stimulation of the cerebral cortex¹, but we know little of the neural mechanisms underlying the cortical modulation of pain. One of the few cortical areas consistently activated by painful stimuli is the rostral agranular insular cortex (RAIC) where, as in other parts of the cortex, the neurotransmitter γ -aminobutyric acid (GABA) robustly inhibits neuronal activity. Here we show that changes in GABA neurotransmission in the RAIC can raise or lower the pain threshold—producing analgesia or hyperalgesia, respectively—in freely moving rats. Locally increasing GABA, by using an enzyme inhibitor or gene transfer mediated by a viral vector, produces lasting analgesia by enhancing the descending inhibition of spinal nociceptive neurons. Selectively activating GABA_B-receptor-bearing RAIC neurons produces hyperalgesia through projections to the amygdala, an area involved in pain and fear. Whereas most studies focus on the role of the cerebral cortex as the end point of nociceptive processing, we suggest that cerebral cortex activity can change the set-point of pain threshold in a top-down manner.

It is well known that painful stimuli activate the insular cortex², but less is known of how the output of the insular cortex might alter pain. Studies in humans and animals show that inhibition or lesion of this cortical area produces analgesia^{3–5}, indicating that the insular cortex tonically produces hyperalgesia. Tonic depression of the nociceptive threshold could result from the activation of pronociceptive areas of the brain or from inhibition of the endogenous pain inhibitory system⁶. To examine the relationship between cortical activity and nociceptive threshold, we manipulated GABA neurotransmission in the RAIC and then tested nociceptive responses with the heat paw-withdrawal test. GABA levels were increased either by inhibiting GABA-aminotransferase-mediated degradation⁷ using vigabatrin, or novel expression of glutamic acid decarboxylase 67 (GAD-67), one of the enzymes responsible for synthesizing GABA from glutamate.

Increasing GABA concentration unilaterally in the RAIC with vigabatrin (78 nmol in 200 nl), resulted in a clear and consistent bilateral analgesia, an effect not seen when the injections were in the surrounding brain areas (Fig. 1a, b). Injection of the GABA_A antagonist bicuculline (200 pmol/200 nl) in the RAIC after vigabatrin normalized the nociceptive threshold, indicating that the effect of vigabatrin was receptor-mediated. Injection of the sodium-channel blocker bupivacaine (200 nl of a 0.25% solution) in the RAIC yielded similar results to vigabatrin (Supplementary Fig. 1a). The antinociceptive effect of vigabatrin seemed specific, because no motor or behavioural impairment was revealed in the open-field (no difference between vigabatrin-treated and saline-treated for any of the parameters; $n = 6$ per group, $P > 0.05$) and rotarod tests (Supplementary Fig. 1b). To determine whether a sustained increase in GABA would produce long-term analgesia, we injected the

defective herpes simplex virus (HSV) vector dvGAD-67 (5×10^3 defective particle units/1.5 μ l) encoding GAD-67. Infection of both neuronal and glial cells with dvGAD-67 leads to a continued synthesis and release of GABA^{8,9} and resulted in bilateral antinociception (Fig. 1c, d) lasting for up to 10 days. Motor and general behaviours were unaltered (Supplementary Fig. 1b–d). Histological analysis confirmed that local expression of the LacZ reporter gene encoding β -Galactosidase (Fig. 1d) was confined to the RAIC and, with the exception of an occasional cell being labelled along the cannula tract, no labelling was found in other brain areas. Injection of a control viral vector expressing only LacZ and alkaline phosphatase had no effect (Fig. 1c), indicating that antinociception did not result from a non-specific event related to the injection of viral vectors.

The above behavioural results demonstrate that GABA-mediated neurotransmission is directly involved in the pain modulatory function of the RAIC. The long duration of the effect obtained with dvGAD-67 suggests that the antinociceptive action occurs through neural mechanisms that do not downregulate over time.

To determine whether the local increase in cortical GABA concentration changed the nociceptive threshold by an action on the descending pain inhibitory system⁶, we administered the non-selective α -adrenoreceptor antagonist phentolamine (3 nmol in 10 μ l) intrathecally over the lumbo-sacral spinal cord through a chronically implanted catheter¹⁰. Phentolamine blocks descending inhibition mediated by noradrenergic bulbo-spinal projections, most of which originate from the locus coeruleus^{11,12}. In all cases phentolamine reversed the antinociception induced by the increased cortical GABA concentration (Fig. 1e), indicating that the antinociceptive effect of inhibiting the RAIC involves an increased activity of noradrenergic bulbo-spinal projections. The dose of phentolamine used here was too low to alter the baseline nociceptive threshold (Fig. 1e).

We then examined projections from the RAIC that might modulate noradrenergic bulbo-spinal neurons. The RAIC was found to have bilateral projections to GAD-immunopositive cells in the caudal brainstem (Fig. 2a–d), including GABAergic cells in nucleus raphe magnus (Fig. 2c, d), many of which project to the locus coeruleus¹³. On the assumption that the output of the RAIC is glutamatergic, the inhibitory effect of RAIC activity on the locus coeruleus is probably mediated by the activation of GABAergic neurons contacted by RAIC afferents (Fig. 2a–d). Indeed, up to two-thirds of afferents onto locus coeruleus neurons are GABAergic¹³ and, together with opioidergic and catecholaminergic afferents, GABAergic afferents constitute the main inhibitory input¹³. It is therefore likely that the analgesia seen after increasing GABA in the RAIC results from a decreased activity of cortical afferents to the peri-coerulear inhibitory neurons, leading to a net increase in the activity of noradrenergic bulbo-spinal neurons (see Fig. 4 for a summary diagram).

GABA acts on neurons through two different types of receptor. In the RAIC, GABA_B receptors are concentrated on pyramidal neurons of cortical layer 5 (Fig. 2e, f), whereas GABA_A receptors are located diffusely throughout all cortical layers on dendrites and soma (Fig. 2g). Because most lamina V neurons are projection neurons, we used retrograde and anterograde tracing to locate their targets. We found that a large contingent of RAIC neurons expressing GABA_B receptors project to the ipsilateral basolateral nucleus of the amygdala ($45.1 \pm 3.9\%$), and few GABA_B neurons project to the peri-coerulear area.

These anatomical findings led us to propose that increasing the GABA concentration in the RAIC and then blocking GABA_B receptors would selectively disinhibit projections to the amygdala, leaving other outputs silent and enabling us to determine the net effect of the disinhibited neurons on pain behaviour. The GABA_B-selective antagonist saclofen (100 nmol/200 nl; see Supplementary

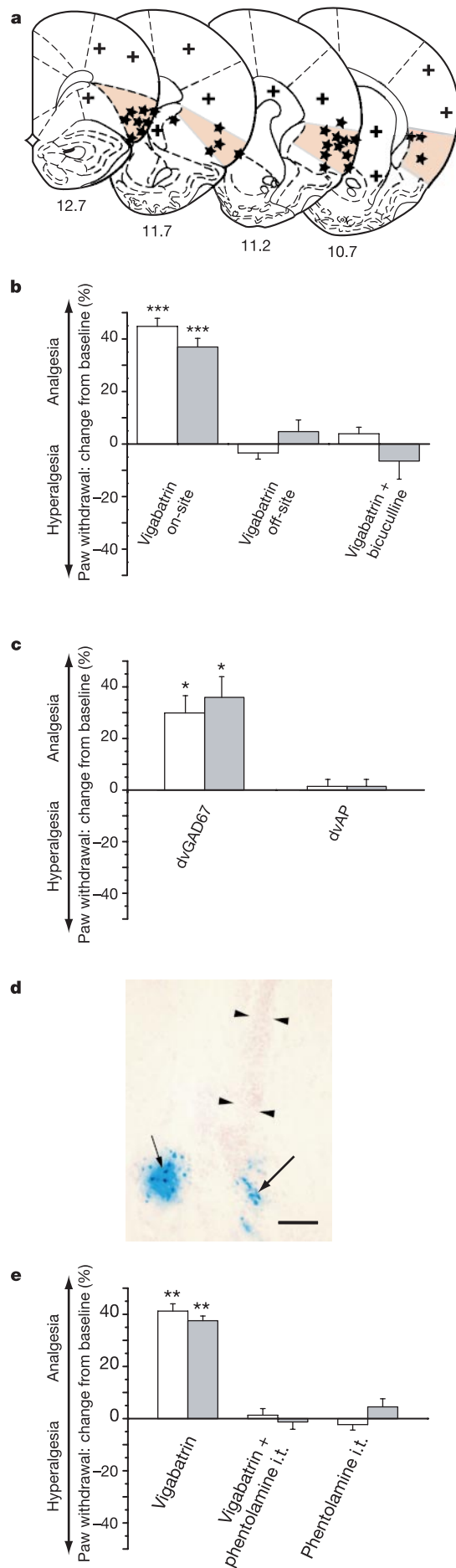


Fig. 2a for a dose–response curve) was microinjected into the RAIC 2 h after vigabatrin or 2–3 days after injection of dvGAD-67; the nociceptive threshold was then assessed. Rats that had become analgesic from vigabatrin or GAD-67 gene expression developed hyperalgesia, principally ipsilateral, immediately after the injection of saclofen (Fig. 3a, b). Fos immunocytochemistry confirmed increased activity in the amygdala (Figs 2h, i and 3c) associated with disinhibition of the GABA_B cortical projection cells and the resulting hyperalgesia. Injection of saclofen alone into the RAIC did not alter the nociceptive threshold (Fig. 3b), indicating that the hyperalgesic effect of disinhibiting GABA_B-receptor-bearing neurons in the RAIC might normally be opposed by other, unknown, neurons bearing GABA_A receptors only.

We then determined that projections from the RAIC to the amygdala, rather than to the caudal brainstem, have a key function in the hyperalgesic state. Bupivacaine (0.25%, 200 nl) injected into the basolateral nucleus of the amygdala after injections of vigabatrin and then saclofen into the RAIC, completely abolished the hyperalgesia, and the nociceptive threshold reverted to the same analgesic state obtained after the injection of vigabatrin alone (Fig. 3a). Ensuing intrathecal injection of phentolamine (3 nmol in 10 μ l) returned the nociceptive threshold from analgesic to normal (Fig. 3a). In contrast, injection of bupivacaine in the locus coeruleus did not alter RAIC induced hyperalgesia (Supplementary Fig. 2b), and blocking the amygdala with bupivacaine had no effect on the analgesia induced by vigabatrin in the RAIC (Fig. 3b).

The RAIC therefore acts on at least two independent subcortical systems to modulate the nociceptive threshold (Fig. 4). The first, from the RAIC to the locus coeruleus, influences the noradrenergic bulbo-spinal projections and is modulated predominantly by GABA_A receptors. The second, from the RAIC to the amygdala, is modulated predominantly by GABA_B receptors. Because GABA_B receptors are metabotropic, their function is probably to keep cortical afferents to the amygdala under prolonged inhibition. However, the effect of GABA_B receptor stimulation is conditional on a number of factors such as the resting potential and concentrations of intracellular chloride¹⁴, factors that significantly influence the outcome of GABA stimulation for cells expressing both types of receptor.

As the nociceptive threshold is reset by blocking RAIC activity with bupivacaine or by locally increasing the GABA concentration, this indicates that both of the above systems might be tonically active. However, tonic GABA release seems to contribute only moderately to the baseline activity of these projections because local injection of bicuculline alone did not significantly alter the nociceptive threshold (Supplementary Fig. 2c). This would imply that a lack of cortical GABA is unlikely to account for states of

Figure 1 Cortical injection sites and nociceptive heat paw-withdrawal responses.

a, Drawings showing the location of the RAIC (shaded area) and representative on-site (asterisks) and off-site (plus signs) injections. Numbers represent rostro-caudal coordinates. **b**, On-site injection of vigabatrin produced a bilateral increase in withdrawal latency, whereas off-site injections had no effect. There was no statistical difference between left (white bars) and right (grey bars) paws. Injection of bicuculline restored the nociceptive threshold to normal. **c**, Injection of GAD-67-expressing vector (dvGAD-67; $n = 6$) produced bilateral analgesia, whereas injection of the control vector expressing alkaline phosphatase (dvAP) produced no behavioural effect. White bars, left paw; grey bars, right paw. **d**, β -Galactosidase-stained section of dvGAD-67-injected RAIC, showing infected cells (arrows) adjacent to the cannula tract (arrowheads). **e**, Analgesia induced by vigabatrin in the RAIC was reversed by intrathecal (i.t.) phentolamine. Phentolamine alone had no effect. One asterisk, $P < 0.05$; two asterisks, $P < 0.01$; three asterisks, $P < 0.001$. Scale bar in **d**, 100 μ m.

hyperalgesia. In fact, a dysfunction of the noradrenergic system, which is a target of the RAIC, probably has a minor function in pain related to chronic inflammation or neuropathic injury, because the complete removal of this system does not, in the long term, alter the nociceptive threshold¹⁵. The direct projections from the RAIC to the caudal brainstem serotonergic cell groups could mediate

some of the pain-modulatory effects of the RAIC, including the hyperalgesia¹⁶.

Our data indicate that if the RAIC is implicated in chronic pain it is through projections to the amygdala, the latter being part of a pro-nociceptive (pain-facilitating) circuit¹⁷. The preferential ipsilateral hyperalgesia after the injection of vigabatrin followed by

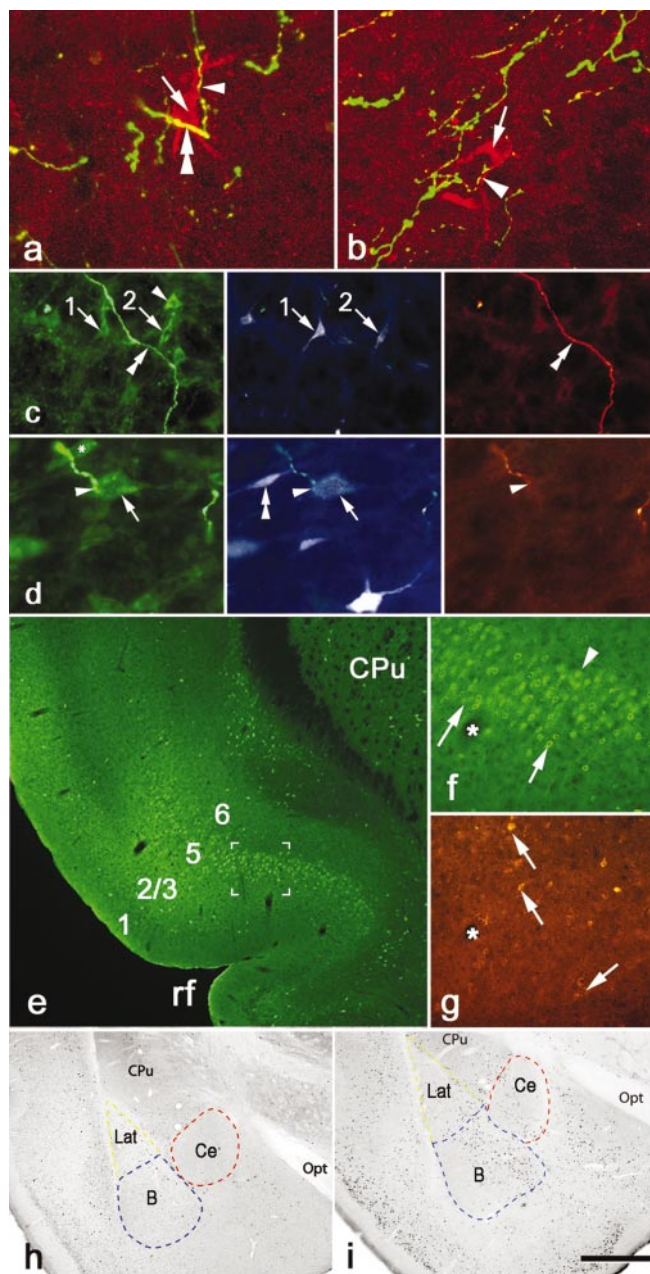


Figure 2 Immunocytochemistry of the RAIC, amygdala and brainstem. **a, b**, Confocal images of GAD-67-immunopositive cells (red, arrows) near the locus coeruleus (**a**) and parabrachial nucleus (**b**) in apposition to labelled axons (green, arrowheads) after injection of biotin-dextran (BDA) into the RAIC. **c**, Nucleus raphe magnus GABAergic neurons (left, arrows 1 and 2) double-labelled with Fluoro-Gold (middle, arrows 1 and 2) from locus coeruleus injection, in apposition to a BDA-labelled fibre from the RAIC (left and right, double arrowhead). Single arrowhead, GABA-only-labelled neuron. **d**, High magnification of a nucleus raphe magnus GABA cell (left, arrow) double-labelled with Fluoro-Gold (middle, arrow) from the locus coeruleus in apposition to puncta from a BDA-labelled fibre (right, arrowhead) from the RAIC. The BDA fibre is also labelled green (left) and is seen faintly as bleed-through fluorescence in the middle panel (arrowhead). Asterisk in left panel, single-labelled GABA-immunoreactive neuron; double arrowhead in

middle panel, single-labelled Fluoro-Gold neuron. **e**, GABA_B-immunopositive pyramidal neurons in lamina 5 of the RAIC. 1–6, cortical layers; rf, rhinal fissure. **f**, GABA_B-immunostained neurons from area in **e** indicated by corners. Outer border (arrows) and inner border (arrowhead) of lamina 5 are indicated. **g**, GABA_A-receptor immunostaining of the same area as **f**. Arrows indicate labelled cell bodies, asterisk indicates the same blood vessel as in **f**. **h**, Fos immunocytochemistry of the amygdala after injection of vigabatrin into the RAIC. **i**, After disinhibition of GABA_B neurons in the RAIC there was an increased number of Fos-immunopositive cells in the basolateral nucleus of the amygdala (B) and to a smaller degree in the central nucleus (Ce). CPu, caudate putamen; Lat, lateral nucleus of the amygdala; Opt, optic tract. Scale bar, 40 μ m for **a, b**; 110 μ m for **c**; 40 μ m for **d**; 350 μ m for **e**; 175 μ m for **f, g**; 1.4 mm for **h, i**.

saclofen into the RAIC is not surprising, given previous reports that the amygdala affects morphine analgesia for the ipsilateral hemi-body¹⁸. The present results indicate that the amygdala can enhance nociceptive responses in addition to its role in opiate analgesia^{18–20}

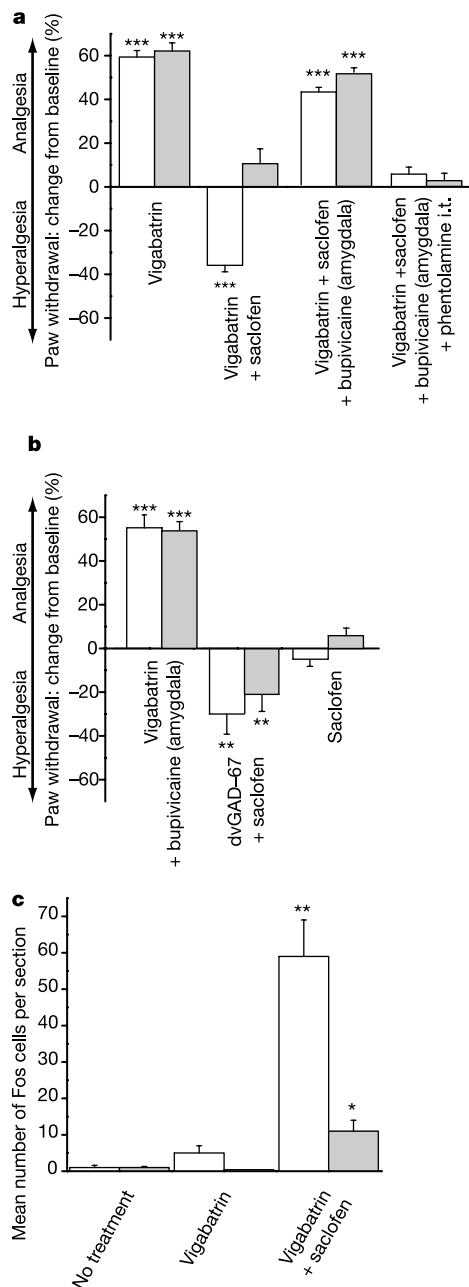


Figure 3 Nociceptive threshold and Fos immunoreactivity after drug treatment. **a**, Changes in nociceptive heat paw-withdrawal responses after the successive injection of four drugs. Each set of bars shows the behavioural response after the injection of the additional drug noted. Vigabatrin and saclofen were injected into the RAIC and the other drugs into the locations indicated. White bars, left paw; grey bars, right paw. **b**, Bupivacaine in the amygdala had no effect on the analgesia produced by vigabatrin alone in the RAIC. Hyperalgesia was also produced by injection of saclofen into the RAIC 2 days after injection of the GAD-67-expressing viral vector dvGAD-67. Injection of saclofen alone into the RAIC did not alter the nociceptive threshold. White bars, left paw; grey bars, right paw. **c**, Number of Fos-immunoreactive cells in the amygdala. White bars, basolateral nucleus; grey bars, central nucleus. One asterisk, $P < 0.05$; two asterisks, $P < 0.01$.

and learned fear²¹. Anaesthesia of the amygdala alone did not alter the nociceptive threshold (Supplementary Fig. 2d), confirming previous reports^{18,21} and showing that the amygdala does not function independently as a 'pain generator' but seems to be controlled directly by cortical afferents. The amygdala is a nodal point where reflexive and learned responses to threatening stimuli are orchestrated. The basolateral nucleus, to which the RAIC has a major projection, is involved in instrumental responses or operant conditioning^{22,23}. In the present protocol, animals were conditioned to interrupt an applied heat stimulus by lifting their paw, a characteristic instrumental response. Rats did not display any signs of stress, such as defensive posture or analgesia encountered in fear reactions, associated with activation of the central nucleus of the amygdala²³. Of the many brain areas to which the basolateral nucleus projects²⁴, the nucleus accumbens is the most likely to initiate a behavioural response to aversive environmental cues^{25,26}. Interestingly, the RAIC has a direct projection to the nucleus accumbens^{27,28}, indicating that interactions between direct and indirect cortical afferents might occur there to modulate nociceptive responses.

Thus, the cerebral cortex modulates pain by acting on both pro-nociceptive and antinociceptive circuits. This dual effect is probably

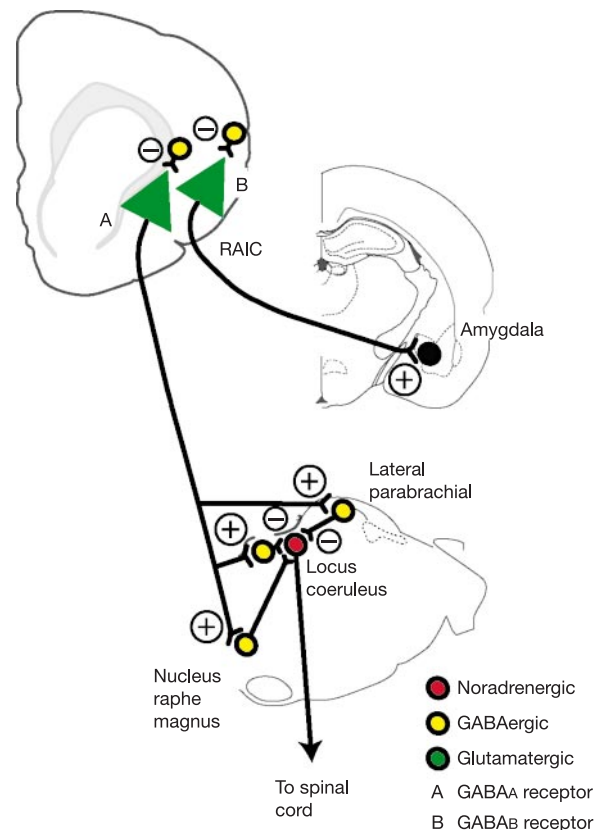


Figure 4 Summary of the main findings of the present study. From the RAIC, at least two independent projections modulate the nociceptive threshold, one to the amygdala and the other to the GABAergic neurons projecting to noradrenergic bulbo-spinal neurons. Both projections are modulated by local cortical GABAergic neurons. GABA_A receptor stimulation (A) predominantly blocks the efferents to the peri-coerulear area, whereas GABA_B receptor stimulation (B) blocks mainly the efferents to the amygdala. At least part of the projection from the RAIC to the peri-coerulear region terminates on GABAergic interneurons, which in turn are presumed to project to noradrenergic neurons in the locus coeruleus.

a defining feature of endogenous pain modulation^{17,21}, and we speculate that an imbalance in the cortical output is likely to underlie some chronic pain states. □

Methods

Experimental animals

We used 252 male Sprague–Dawley rats (270–320 g; Bantin-Kingman). Procedures for the maintenance and use of the experimental animals conformed to the regulations of the UCSF Committee on Animal Research.

Implantation of intracerebral cannulae

A stainless steel guide cannula (26-gauge, Plastics One) was cemented over a burrhole drilled over the RAIC (anterior-posterior (AP) 11.0, lateral (lat) 3.5 mm), and/or the amygdala (AP 6.20, lat 4.8 mm), and/or the locus coeruleus (AP −0.80, lat 1.2 mm), with interaural zero as the reference point. The guide cannula did not extend below the bone, to avoid any damage to the brain. Cortical injections were unilateral because stimulation of the RAIC with drugs was previously shown to produce bilateral effects⁴. On the day of testing (10 days after surgery), a 33-gauge bevelled injection cannula was inserted through the guide cannula to a distance 5.8, 8.0 or 6.0 mm below the cortical surface for the RAIC, amygdala or locus coeruleus, respectively. Drugs and viral vectors were injected over a 1-min period with a microinfusion pump.

Intracerebral injections of tracers

Fluoro-Gold (4%; 10–40 nl) in double-distilled water or 10% biotin–dextran (Molecular Probes) in PBS pH 7.4 was injected through a micropipette (40-μm tip). A period of 2–4 days was allowed for tracer transport.

Nociceptive testing

A treatment-blind researcher conducted the behavioural experiments. Hindpaw responses to radiant heat were measured with a commercial heat paw-withdrawal device (Plantar Analgesia Instrument; Ugo Basile, Comerio, Italy) in accordance with a standard protocol¹⁵.

Open-field test

The open-field test for motor and behavioural impairment was performed in accordance with a standard protocol¹⁵.

Histology

Aldehyde-fixed brains were cut transversely (40 μm slices) and kept in rostro-caudal order. Intracerebral drug injection sites were mapped on serial Nissl-stained sections. Histochemistry for β-galactosidase and acid phosphatase²⁹, and immunocytochemistry, were performed in accordance with standard protocols. Guinea-pig anti-GABA_A antiserum was provided by Dr Margeta-Mitrovic (Departments of Physiology and Biochemistry, University of California, San Francisco)³⁰, and rabbit anti-fofos antiserum was provided by Dr. D. Slamon (Department of Medicine, University of California, Los Angeles). Rabbit anti-GAD-67 (Ab5862) antiserum was purchased from Chemicon, and mouse monoclonal anti-GABAA (05-474) from Upstate Biotechnology.

Stereology and analysis of immunolabelled cells

Counts of cells per unit area of the cortex or amygdala were performed with an unbiased counting method with the use of a computer-aided system (Stereo Investigator; MicroBrightField).

Drugs

Vigabatrin, saclofen (Tocris Cookson), phentolamine (Sigma-Aldrich) and bupivacaine (AstraZeneca) were dissolved in 0.9% saline.

Viral vectors

Double-cassette-defective HSV vectors (dvGAD-67 and dvAP) were generated from amplicon plasmids, pHCL-CGAD-67 encoding rat GAD-67 (obtained from A. Tobin, University of California, Los Angeles) and *Escherichia coli* LacZ, and pHCL-CAP encoding *E. coli* LacZ and human acid phosphatase²⁹, with HSV-1 tsK as helper virus⁸. The titre of the defective vectors (in defective particle units) was determined by histochemistry with 5-bromo-4-chloro-3-indolyl-β-D-galactoside. *In vitro*, infection of cultured cerebellar granule cells (CGCs)—predominantly glutamatergic—or of cultured astrocytes resulted in both the expression of GAD-67 and the synthesis of GABA. GABA is released tonically from dvGAD-67-infected astrocytes and in a stimulus-evoked fashion from CGCs^{8,9}. Expression from these vectors, with the CMV_{IE} promoter/enhancer, continues for about 7–10 days *in vivo*. Control rats were injected with a defective HSV vector expressing alkaline phosphatase and LacZ.

Statistical analysis

By using the outcome variable in each experiment, comparisons of treatment groups were made with Student's *t*-test or analysis of variance; *P* < 0.05 was considered statistically significant. Post-hoc analysis of means and s.e.m. with Scheffé's *F*-test confirmed statistical significance.

Received 7 January; accepted 28 May 2003; doi:10.1038/nature01808.

1. Calejesan, A. A., Kim, S. J. & Zhuo, M. Descending facilitatory modulation of a behavioral

- nociceptive response by stimulation in the adult rat anterior cingulate cortex. *Eur. J. Pain* **4**, 83–96 (2000).
- Ostrowsky, K. *et al.* Representation of pain and somatic sensation in the human insula: A study of responses to direct electrical cortical stimulation. *Cereb. Cortex* **12**, 376–385 (2002).
- Berthier, M., Starkstein, S. & Leiguarda, R. Asymbolia for pain: A sensory-limbic disconnection syndrome. *Ann. Neurol.* **24**, 41–49 (1988).
- Burkey, A. R., Carstens, E. & Jasmin, L. Dopamine reuptake inhibition in the rostral agranular insular cortex produces antinociception. *J. Neurosci.* **19**, 4169–4179 (1999).
- Greenspan, J. D. & Winfield, J. A. Reversible pain and tactile deficits associated with a cerebral tumor compressing the posterior insula and parietal operculum. *Pain* **50**, 29–39 (1992).
- Hunt, S. P. & Mantyh, P. W. The molecular dynamics of pain control. *Nature Rev. Neurosci.* **2**, 83–91 (2001).
- Jung, M. J., Lippert, B., Metcalf, B. W., Bohlen, P. & Schechter, P. J. γ-Vinyl GABA (4-amino-hex-5-enoic acid), a new selective irreversible inhibitor of GABA-T: Effects on brain GABA metabolism in mice. *J. Neurochem.* **29**, 797–802 (1977).
- New, K. C., Gale, K., Martuza, R. L. & Rabkin, S. D. Novel synthesis and release of GABA in cerebellar granule cell cultures after infection with defective herpes simplex virus vectors expressing glutamic acid decarboxylase. *Brain Res. Mol. Brain Res.* **61**, 121–135 (1998).
- New, K. C. & Rabkin, S. D. GABA synthesis in astrocytes after infection with defective herpes simplex virus vectors expressing glutamic acid decarboxylase 65 or 67. *J. Neurochem.* **71**, 2304–2312 (1998).
- Jasmin, L. & Ohara, P. T. Long-term intrathecal catheterization in the rat. *J. Neurosci. Methods* **110**, 81–89 (2001).
- Proudfit, H. K. & Clark, F. M. The projections of locus coeruleus neurons to the spinal cord. *Prog. Brain Res.* **88**, 123–141 (1991).
- Sagen, J. & Proudfit, H. K. Effect of intrathecally administered noradrenergic antagonists on nociception in the rat. *Brain Res.* **310**, 295–301 (1984).
- Somogyi, J. & Llewellyn-Smith, I. J. Patterns of colocalization of GABA, glutamate and glycine immunoreactivities in terminals that synapse on dendrites of noradrenergic neurons in rat locus coeruleus. *Eur. J. Neurosci.* **14**, 219–228 (2001).
- Lopantsev, V. & Schwartzkroin, P. A. GABA_A-dependent chloride influx modulates GABA_B-mediated IPSPs in hippocampal pyramidal cells. *J. Neurophysiol.* **82**, 1218–1223 (1999).
- Jasmin, L., Boudah, A. & Ohara, P. T. Long-term effects of decreased noradrenergic central nervous system innervation on pain behavior and opioid antinociception. *J. Comp. Neurol.* **460**, 38–55 (2003).
- Suzuki, R., Morcuende, S., Webber, M., Hunt, S. P. & Dickenson, A. H. Superficial NK1-expressing neurons control spinal excitability through activation of descending pathways. *Nature Neurosci.* **5**, 1319–1326 (2002).
- Fields, H. L. Pain modulation: expectation, opioid analgesia and virtual pain. *Prog. Brain Res.* **122**, 245–253 (2000).
- Manning, B. H. A lateralized deficit in morphine antinociception after unilateral inactivation of the central amygdala. *J. Neurosci.* **18**, 9453–9470 (1998).
- Helmstetter, F. J., Tershner, S. A., Poore, L. H. & Bellgowan, P. S. Antinociception following opioid stimulation of the basolateral amygdala is expressed through the periaqueductal gray and rostral ventromedial medulla. *Brain Res.* **779**, 104–118 (1998).
- Nandigama, P. & Borszcz, G. S. Affective analgesia following the administration of morphine into the amygdala of rats. *Brain Res.* **959**, 343–354 (2003).
- Watkins, L. R. *et al.* Neurocircuitry of conditioned inhibition of analgesia: Effects of amygdala, dorsal raphe, ventral medullary, and spinal cord lesions on antianalgesia in the rat. *Behav. Neurosci.* **112**, 360–378 (1998).
- Killcross, S., Robbins, T. W. & Everitt, B. J. Different types of fear-conditioned behaviour mediated by separate nuclei within amygdala. *Nature* **388**, 377–380 (1997).
- LeDoux, J. in *The Amygdala* (ed. Aggleton, J.) 289–310 (Oxford Univ. Press, 2000).
- Pitkanen, A. in *The Amygdala* (ed. Aggleton, J.) 31–115 (Oxford Univ. Press, 2000).
- Zhang, S., Tang, J. S., Yuan, B. & Jia, H. Involvement of the frontal ventrolateral orbital cortex in descending inhibition of nociception mediated by the periaqueductal gray in rats. *Neurosci. Lett.* **224**, 142–146 (1997).
- Everitt, B. J. & Robbins, T. W. in *The Amygdala* (ed. Aggleton, J.) 401–429 (Wiley-Liss, New York, 1992).
- Brog, J. S., Salyapongse, A., Deutch, A. Y. & Zahm, D. S. The patterns of afferent innervation of the core and shell in the 'accumbens' part of the rat ventral striatum: Immunohistochemical detection of retrogradely transported fluoro-gold. *J. Comp. Neurol.* **338**, 255–278 (1993).
- Berendse, H. W., Galis-de Graaf, Y. & Groenewegen, H. J. Topographical organization and relationship with ventral striatal compartments of prefrontal corticostriatal projections in the rat. *J. Comp. Neurol.* **316**, 314–347 (1992).
- New, K. C. & Rabkin, S. D. Co-expression of two gene products in the CNS using double-cassette defective herpes simplex virus vectors. *Brain Res. Mol. Brain Res.* **37**, 317–323 (1996).
- Margeta-Mitrovic, M., Mitrovic, I., Riley, R. C., Jan, L. Y. & Basbaum, A. I. Immunohistochemical localization of GABA_B receptors in the rat central nervous system. *J. Comp. Neurol.* **405**, 299–321 (1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. D. Levine for critical comments, H. J. Ralston for support and encouragement, K. New for constructing the viral vector, U. MacGarvey for assistance with histochemistry, and G. Janni for editorial assistance. This work was funded by NIH (L.J.), NINDS (S.D.R.) and the Koret Foundation (P.T.O.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.J. (ucpain@itsa.ucsf.edu) or P.T.O. (pto@itsa.ucsf.edu).

A cytoplasmic region determines single-channel conductance in 5-HT₃ receptors

Stephen P. Kelley*, James I. Dunlop*, Ewen F. Kirkness†, Jeremy J. Lambert* & John A. Peters*

* Neurosciences Institute, Department of Pharmacology and Neuroscience, Ninewells Hospital and Medical School, The University of Dundee, Dundee DD1 9SY, UK

† The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, Maryland 20850, USA

5-Hydroxytryptamine type 3 (5-HT₃) receptors are cation-selective transmitter-gated ion channels of the Cys-loop superfamily^{1–9}. The single-channel conductance of human recombinant 5-HT₃ receptors assembled as homomers of 5-HT_{3A} subunits, or heteromers of 5-HT_{3A} and 5-HT_{3B} subunits, are markedly different, being 0.4 pS (refs 6, 9) and 16 pS (ref. 7), respectively. Paradoxically, the channel-lining M2 domain of the 5-HT_{3A} subunit would be predicted to promote cation conduction, whereas that of the 5-HT_{3B} subunit would not⁷. Here we describe a determinant of single-channel conductance that can explain these observations. By constructing chimaeric 5-HT_{3A} and 5-HT_{3B} subunits we identified a region (the 'HA-stretch')¹⁰ within the large cytoplasmic loop of the receptor that markedly influences channel conductance. Replacement of three arginine residues unique to the HA-stretch of the 5-HT_{3A} subunit by their

5-HT_{3B} subunit counterparts increased single-channel conductance 28-fold. Significantly, ultrastructural studies of the *Torpedo* nicotinic acetylcholine receptor¹¹ indicate that the key residues might frame narrow openings that contribute to the permeation pathway. Our findings solve the conundrum of the anomalously low conductance of homomeric 5-HT_{3A} receptors and indicate an important function for the HA-stretch in Cys-loop transmitter-gated ion channels.

To identify the 5-HT_{3B} subunit amino-acid sequences responsible for the greatly enhanced single-channel conductance (γ) of heteromeric (5-HT_{3A} and 5-HT_{3B}) over homomeric (5-HT_{3A}) receptors, we generated 10 chimaeric constructs (C1–C10; Fig. 1a) of the human 5-HT_{3A} and 5-HT_{3B} subunits. These were expressed, singly or in combination, with the wild-type 5-HT_{3A} subunit in tsA-201 cells. In addition, the wild-type subunits and chimaeric constructs incorporated small haemagglutinin or Myc epitopes inserted at the extreme amino terminus (between residues 5 and 6 of the mature protein) to assess their cellular distribution in COS 7 cells by immunofluorescence under confocal microscopy¹². We have previously shown that the insertion of such epitope tags is functionally silent¹². Fluctuation analysis of 5-HT-induced whole-cell currents recorded under voltage clamp^{6,7} was used to compare the single-channel conductance of the subunit assemblies because the sub-picoampere conductance of the homomeric 5-HT_{3A} receptor precludes the direct resolution of single-channel events^{6,7,13,14}.

Initial experiments (Fig. 1b, c) verified the large difference in γ between homomeric (0.76 pS) and heteromeric (13.3 pS) 5-HT₃ receptors and confirmed, for the latter, that fluctuation analysis yields a γ comparable to that obtained by single-channel recording (16 pS) (ref. 7). The chimaeric subunits were subsequently analysed for the formation of functional receptors when expressed alone, and

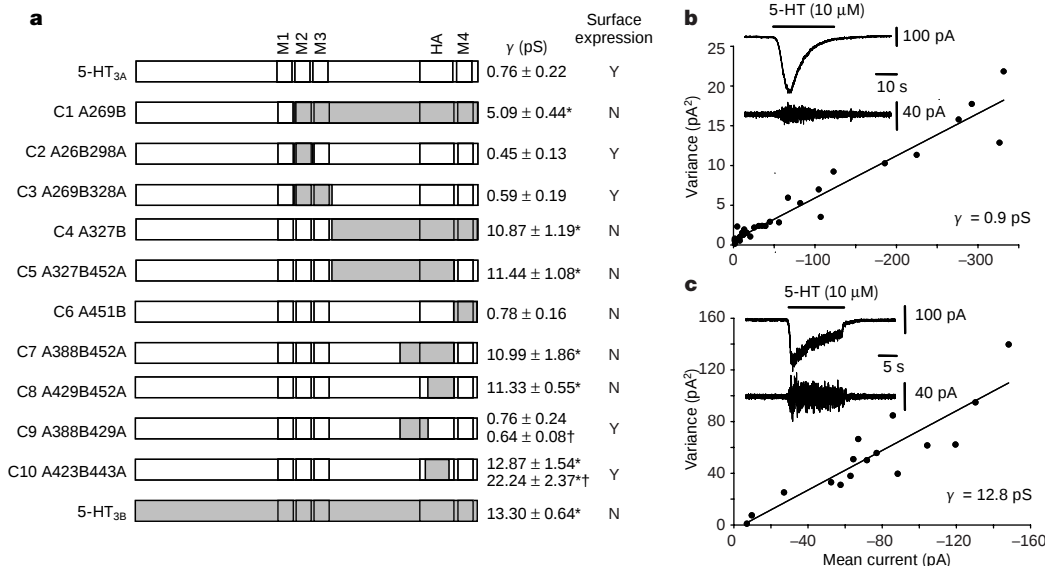


Figure 1 The strategy employed to identify the domain of the 5-HT_{3B} subunit conferring enhanced single-channel conductance on heteromeric (5-HT_{3A} and 5-HT_{3B}) and on homomeric (5-HT_{3A}) receptors. **a**, Schematic depiction of the 5-HT_{3A}/5-HT_{3B} subunit chimaeras C1–C10 defined according to the following convention: A indicates the N-terminal 5-HT_{3A} subunit sequence and the following number indicates the position of the last 5-HT_{3A} residue before a transition to 5-HT_{3B} subunit sequence (denoted B); where appropriate the following number indicates the first residue in a resumption to 5-HT_{3A} subunit sequence. M1, M2, M3, M4 and HA indicate the positions of the transmembrane domains and HA-stretch. Single-channel conductances (γ) are summarized in the first

right-hand column. A dagger indicates γ for C9 and C10 expressed as homomers. Data are means ± s.e.m. from 5–11 cells. An asterisk indicates a significant difference ($P < 0.05$) from the wild-type homomeric receptor (one-way analysis of variance and subsequent *post hoc* tests). Cell-surface expression in the absence of the 5-HT_{3A} subunit is summarized in the second right-hand column (Y, yes; N, no). **b**, **c**, Representative low-gain (direct current)-coupled and high-gain (alternating current)-coupled records of inward current responses to 5-HT (10 μ M) recorded from tsA-201 cells transfected with 5-HT_{3A} cDNA (**b**) or 5-HT_{3A} and 5-HT_{3B} cDNA (**c**), and the corresponding current variance plots.

for their influence on single-channel conductance when co-expressed with the 5-HT_{3A} subunit. Transfection of tsA-201 cells with complementary DNAs encoding chimaeras C1 and C4–C8, in the absence of co-expressed 5-HT_{3A} subunits, did not support inward current responses to 5-HT (10 μ M) and cell-surface expression could not be detected by immunofluorescence, although intracellular staining was prominent in permeabilized cells. However, when co-expressed with the 5-HT_{3A} subunit, the formation of

heteromeric complexes between 5-HT_{3A} and chimaeric subunits was demonstrated by cell-surface labelling and, for chimaeras C1, 4, 5, 7 and 8, a large enhancement of γ (7.6–15-fold) over that observed for the homomeric 5-HT_{3A} receptor (Fig. 1a). Although chimaeras C2 and C3 did traffic to the cell surface in the absence of the 5-HT_{3A} subunit, no current response to 5-HT could be detected. It is possible that the putative channel formed by these chimaeras is not conducive to ion permeation, or does not gate in response to agonist binding.

The value of γ (5.1 pS) demonstrated by co-expressed C1 and 5-HT_{3A} subunits provided the first indication that 5-HT_{3B} subunit sequence carboxy-terminal to the intracellular border of M2 (Fig. 1a) contained a domain responsible for enhanced γ . Significantly, the results obtained with C2 and C3 (Fig. 1a) exclude M2 (and M3) in this effect, assuming that they assemble with the 5-HT_{3A} subunit. This seems reasonable, because C1, embodying both M2 and M3 of the 5-HT_{3B} subunit sequence, clearly assembles with the 5-HT_{3A} subunit. C4, incorporating both the large intracellular loop and M4 of the 5-HT_{3B} subunit sequence, enhanced γ to a value (10.9 pS) close to that observed for wild-type heteromeric 5-HT₃ receptors. The data obtained with C5 and C6 (Fig. 1a) indicate the intracellular loop to be the locus of this effect. Progressive fractionation of C5, by way of C7, C8 and C9, identified a stretch of 23 amino acids containing a predicted amphipathic α -helix, within the 'HA-stretch' in nicotinic acetylcholine (nACh) receptor subunits^{10,11}, to be sufficient to increase γ to 11.3 pS, which is not significantly different from that of the wild-type heteromer (Fig. 1a). Further restricting the substituted region (C10; Figs 1a and 2) was without effect on the enhancement of γ (12.9 pS). Moreover, C10 assembled into functional homomeric receptors, yielding a γ (22.2 pS) significantly higher than that obtained when co-expressed with the 5-HT_{3A} subunit (Fig. 2a, b). C9, containing 23 5-HT_{3B} subunit residues immediately N-terminal to C10, also functioned as a homomer, but with a γ (0.64 pS) indistinguishable from that of homomeric 5-HT_{3A} receptors.

An alignment of the HA-stretch across species orthologues of 5-HT₃ receptor subunits (Fig. 2c) and representative subunits of cation-selective Cys-loop transmitter-gated ion channels (Fig. 2d) reveals the presence of four positively charged arginine residues unique to the 5-HT_{3A} subunit sequence. We therefore mutated these residues, individually and in combination, to the aligned residues in the human 5-HT_{3B} subunit sequence and expressed the resulting receptors as homomers (Fig. 3a). The mutations R426S and R432Q had little effect on γ , whereas R436D and R440A caused ~5-fold and ~3-fold increases in γ respectively (Fig. 3a). The double mutations R432Q/R436D and R436D/R440A produced enhancements of γ that were greater than the sum of the individual mutations, but a similar synergy was not observed for R432Q/R440A (Fig. 3a). That the major determinants of γ are indeed the

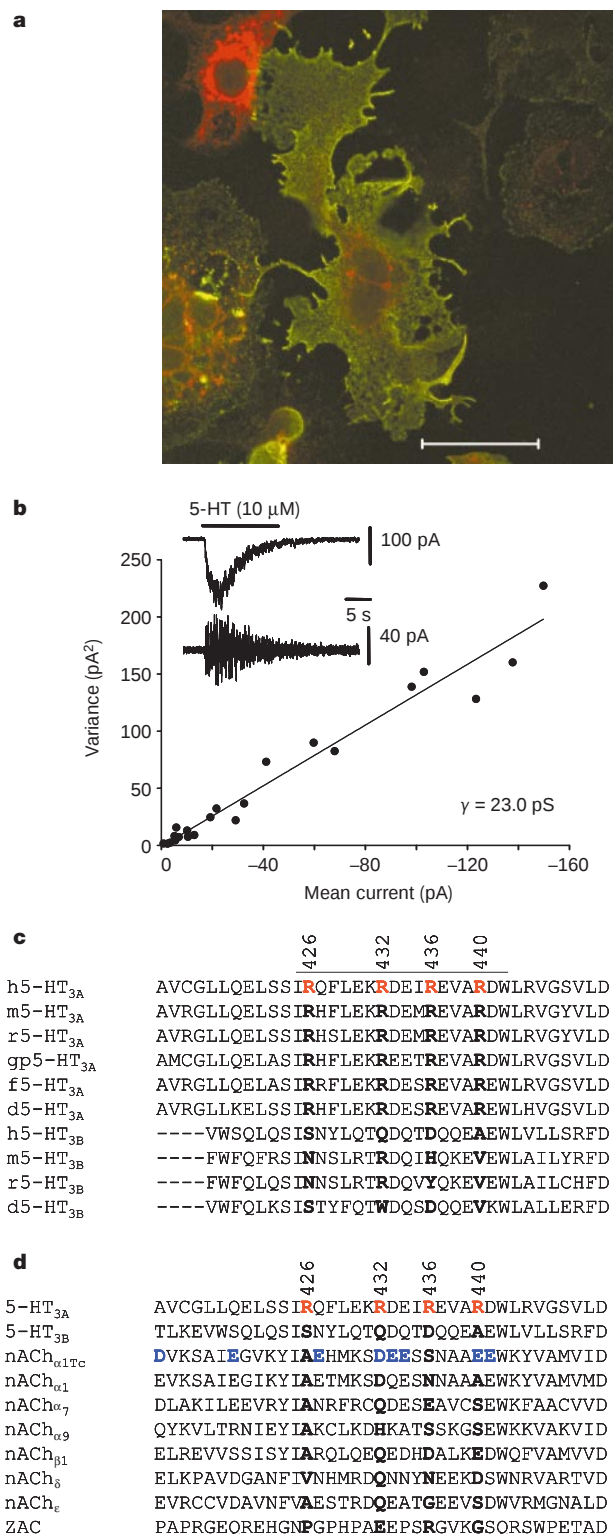


Figure 2 Identification of the domain of the 5-HT_{3B} subunit responsible for the enhanced conductance of hetero-oligomeric compared with homo-oligomeric 5-HT₃ receptors.

a, Photomicrograph of COS 7 cells expressing C10. Cell-surface and intracellular expression are indicated by green and red fluorophores, respectively. Scale bar, 50 μ m. **b**, Representative low-gain (direct current)-coupled and high-gain (alternating current)-coupled records of current responses to 5-HT (10 μ M) recorded from tsA-201 cells transfected with C10 cDNA, and the corresponding current variance plot. **c**, Alignment of the HA-stretch across 5-HT_{3A} and 5-HT_{3B} subunit species orthologues. The bar indicates the location of C10 residues; red and blue letters indicate residues subjected to mutagenesis in the human 5-HT_{3A} and *Torpedo* nACh α 1 subunits⁸, respectively. h, human; m, mouse; r, rat; gp, guinea-pig; f, ferret; d, dog (E.F.K., unpublished observations). **d**, Alignment of the HA-stretch across representative subunits of human cationic transmitter-gated ion channels and the α 1 subunit of the nACh receptor of *Torpedo californica* (Tc). ZAC, zinc-activated channel²⁹.

unique arginine residues is apparent from the triple mutation R432Q/R436D/R440A, which yielded a γ (21.5 pS) not significantly different from that of C10 incorporating all of the residues within this region. The greatly enhanced γ of the R432Q/R436D/R440A mutant was confirmed by recordings of 5-HT-induced single-channel currents from outside-out membrane patches, which provided an estimate of γ of 25.1 ± 0.8 pS ($n = 3$) at -60 mV (Fig. 3a, b). It is unlikely that the triple mutation introduced a gross change in ion selectivity, because the reversal potential of 5-HT-induced whole-cell currents (-3.1 ± 0.8 mV, $n = 4$) was similar to that of the 5-HT_{3A} receptor (-1.7 ± 1.0 mV, $n = 3$) and was unchanged when intracellular chloride ions were partly replaced by gluconate anions (-2.9 ± 2.6 mV, $n = 3$). Finally, reversal of the sign of the charges presented by Arg 432, Arg 436 and Arg 440, by their collective mutation to glutamate, did not

provide any further enhancement of γ as determined by fluctuation analysis (18.9 ± 3.0 pS; $n = 6$; not shown).

Our data identify uniquely occurring arginine residues within the 5-HT_{3A} subunit as prominent influences on γ . Previously, the HA-stretch has been shown to influence the gating kinetics^{15,16} and assembly¹⁷ of nACh receptors, but not γ . No significant change in γ was noted¹⁸ when acidic residues within the HA-stretch of the *Torpedo* nACh α -subunit were mutated singly (or in one case doubly) to lysine (residues in blue in Fig. 2d). However, only two mutations were at positions homologous to those studied here (Fig. 2d) and, considering stoichiometry, apply to two, not five, subunits¹⁸. There is strong evidence that the M2 segments of 5-HT_{3A} subunits contribute to the pore of the channel^{19–21}. We speculate that the pore funnels into a vestibule perforated by openings ('portals') through which ions flux to the cytoplasm and vice versa (Fig. 3c). Such cytoplasmic portals exist in the *Torpedo* nACh receptor¹¹, formed between adjacent subunits by α -helices identifiable with the HA-stretch^{11,22}. The dimensions of the portals are estimated to be no greater than the diameter of a permeating ion, including its first hydration shell^{11,22}. Thus, the charge of the residues framing the portals should affect both ionic movements and the local concentration of ions within the vestibule²². We propose that the anomalously low conductance of the homomeric 5-HT_{3A} receptor is attributable to an unfavourable charge distribution over the portals (or at least in proximity to them) provided by Arg 432, Arg 436 and Arg 440. Presumably, five identical portals exist in the inner vestibule of the homomeric 5-HT_{3A} receptor, allowing speculation that the 5-HT_{3B} subunit increases γ by the formation of cytoplasmic 5-HT_{3A}/5-HT_{3B} subunit interfaces that are more conducive to cation conduction.

The existence of portals in the cytoplasmic protrusion of ion-channel proteins is a recurring structural theme. Similar structures (within 'hanging gondolas')²³ have been described for the Shaker potassium channel^{23,24}, the bacterial mechanosensitive channel MscL²⁵, the voltage-activated sodium channel²⁶ and a cyclic-GMP-gated channel²⁷. Our assignment of an important function to the inner vestibule in ion permeation might therefore have broad significance for ion-channel function. □

Methods

DNA constructs and transient transfection of subunit cDNAs

Human 5-HT_{3A} and 5-HT_{3B} subunit cDNAs (ref. 3) and cDNAs encoding chimaeric constructs of the 5-HT_{3A/3B} subunits were each cloned into pGW1 (ref. 12). Chimaeric 5-HT_{3A/3B} sequences were generated by standard molecular biological techniques²⁸. The subunits were tagged with the Myc or haemagglutinin epitopes between amino acids 5 and 6 of the mature protein by site-directed mutagenesis, as described previously¹². All cDNAs were completely sequenced to confirm fidelity. Transfection of tsA-201 or COS-7 cells with subunit cDNAs, at equimolar ratios when appropriate, was performed by electroporation (400 V, infinite resistance, 125 μ F) with the use of a Bio-Rad gene electropulser II. Transfected cells were cultured for 24–72 h before use.

Immunofluorescence

Anti-Myc and anti-haemagglutinin antibodies were obtained from 9E10 or 12CA5 hybridoma cells, respectively, and used direct as supernatant (20 μ g ml⁻¹). The fluorophore-conjugated secondary antibodies goat anti-mouse Alexa Fluor 488 (green) and Alexa Fluor 568 (red) (Molecular Probes) were used at 1 μ g ml⁻¹. Immunofluorescence was performed as described previously¹² with a confocal microscope (Zeiss LSM510).

Electrophysiology

Whole-cell and outside-out patch configurations were used to record macroscopic and single-channel currents respectively from transfected tsA-201 cells. The recording chamber was perfused (5 ml min⁻¹) with a solution of (in mM): NaCl 140, KCl 2.8, MgCl₂ 2.0, CaCl₂ 1.0, glucose 10, HEPES 10, pH 7.3. Patch-clamp electrodes contained (in mM): KCl 140, MgCl₂ 2.0, CaCl₂ 0.1, EGTA 1.1, HEPES 10, pH 7.3. In experiments examining ionic selectivity, [Cl⁻] within the pipette solution was decreased to 15 mM by the partial replacement of KCl by potassium gluconate. 5-HT was applied to voltage-clamped cells (-60 mV) with a diffusion pipette positioned to elicit relatively slowly rising and decaying currents suited to fluctuation analysis. Single-channel conductances were estimated by fluctuation analysis (1–1000-Hz bandwidth) of the macroscopic current response to 10 μ M 5-HT recorded at a holding potential of -60 mV as described previously⁶, or by direct observation of single-channel currents (filtered at 2 kHz, digitized at 20 kHz)

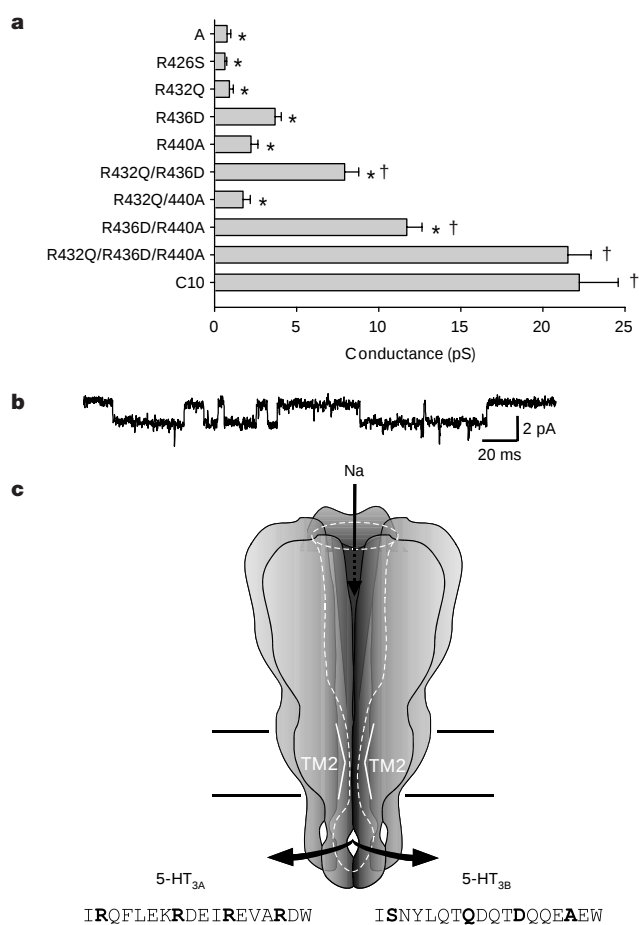


Figure 3 Conserved arginine residues within the HA-stretch of the 5-HT_{3A} receptor subunit influence single-channel conductance. **a**, Bar plot comparing the single-channel conductance of homo-oligomeric receptors constructed from wild-type 5-HT_{3A} subunits (A), or C10, with receptors assembled from 5-HT_{3A} subunits harbouring single, or multiple, mutations. Data are means $\pm$ s.e.m. from five to nine cells. There was an overall significant effect of construct on channel conductance ($F_{9,61} = 55.2$, $P < 0.001$). A dagger indicates a significant difference from the wild-type 5-HT_{3A} receptor; an asterisk indicates a significant difference from the receptor assembled from C10 (Tukey's t , $P < 0.05$). **b**, Single-channel currents evoked by 5-HT (10 μ M) applied to an outside-out membrane patch (holding potential -60 mV) excised from a tsA-201 cell transfected with the 5-HT_{3A} subunit mutant R432Q/R436D/R440A. **c**, Schematic depiction of the 5-HT₃ receptor indicating the putative cytoplasmic portals. Human 5-HT_{3A} and 5-HT_{3B} subunit 'HA-stretch' sequences proposed to contribute to the portals are illustrated^{11,22}.

elicited by 10 μ M 5-HT applied to outside-out membrane patches. Data analysis was performed with the Electrophysiology Data Recorder developed and provided by J. Dempster (Department of Physiology and Pharmacology, University of Strathclyde, UK; <http://www.strath.ac.uk/Departments/PhysPharm/>). Data are reported as means $\pm$ s.e.m.

Received 2 April; accepted 20 May 2003; doi:10.1038/nature01788.

1. Derkach, V., Surprenant, A. & North, R. A. 5-HT₃ receptors are membrane ion channels. *Nature* **339**, 706–709 (1989).
2. Yakel, J. L. & Jackson, M. B. 5-HT₃ receptors mediate rapid responses in cultured hippocampus and a clonal cell line. *Neuron* **1**, 615–621 (1988).
3. Yakel, J. L., Shao, X. M. & Jackson, M. B. The selectivity of the channel coupled to the 5-HT₃ receptor. *Brain Res.* **533**, 46–52 (1990).
4. Yang, J. Ion permeation through 5-hydroxytryptamine-gated channels in neuroblastoma N18 cells. *J. Gen. Physiol.* **96**, 1177–1198 (1990).
5. Maricq, A. V., Peterson, A. S., Brake, A. J., Myers, R. M. & Julius, D. Primary structure and functional expression of the 5-HT₃ receptor, a serotonin-gated ion channel. *Science* **254**, 432–437 (1991).
6. Brown, A. M., Hope, A. G., Lambert, J. J. & Peters, J. A. Ion permeation and conduction in a human recombinant 5-HT₃ receptor subunit (h5-HT_{3A}). *J. Physiol. (Lond.)* **507**, 653–665 (1998).
7. Davies, P. A. *et al.* The 5-HT_{3B} subunit is a major determinant of serotonin-receptor function. *Nature* **397**, 359–363 (1999).
8. Dubin, A. E. *et al.* The pharmacological and functional characteristics of the serotonin 5-HT_{3A} receptor are specifically modified by a 5-HT_{3B} receptor subunit. *J. Biol. Chem.* **274**, 30799–30810 (1999).
9. Hussy, N., Lukas, W. & Jones, K. A. Functional properties of a cloned 5-hydroxytryptamine ionotropic receptor subunit: Comparison with native mouse receptors. *J. Physiol. (Lond.)* **481**, 311–323 (1994).
10. Finer-Moore, J. & Stroud, R. M. Amphipathic analysis and possible formation of the ion channel in an acetylcholine receptor. *Proc. Natl Acad. Sci. USA* **81**, 155–159 (1984).
11. Miyazawa, A., Fujiyoshi, Y., Stowell, M. & Unwin, N. Nicotinic acetylcholine receptor at 4.6 Å resolution: Transverse tunnels in the channel wall. *J. Mol. Biol.* **288**, 765–786 (1999).
12. Boyd, G. W. *et al.* Assembly and cell surface expression of homomeric and heteromeric 5-HT₃ receptors: The role of oligomerization and chaperone proteins. *Mol. Cell. Neurosci.* **21**, 38–50 (2002).
13. Gill, C. H., Peters, J. A. & Lambert, J. J. An electrophysiological investigation of the properties of a murine recombinant 5-HT₃ receptor stably expressed in HEK 293 cells. *Br. J. Pharmacol.* **114**, 1211–1221 (1995).
14. Hanna, M. C., Davies, P. A., Hales, T. G. & Kirkness, E. F. Evidence for expression of heteromeric serotonin 5-HT₃ receptors in rodents. *J. Neurochem.* **75**, 240–247 (2002).
15. Bouzat, C., Bren, N. & Sine, S. M. Structural basis of the different gating kinetics of fetal and adult acetylcholine receptors. *Neuron* **13**, 1395–1402 (1994).
16. Akk, G. & Steinbach, J. H. Structural elements near the C-terminus are responsible for changes in nicotinic receptor gating kinetics following patch excision. *J. Physiol. (Lond.)* **527**, 405–417 (2000).
17. Yu, X. M. & Hall, Z. W. A sequence in the main cytoplasmic loop of the alpha subunit is required for assembly of mouse muscle nicotinic acetylcholine receptor. *Neuron* **13**, 247–255 (1994).
18. Imoto, K. *et al.* Rings of negatively charged amino acids determine the acetylcholine receptor channel conductance. *Nature* **335**, 645–648 (1988).
19. Reeves, D. C., Goren, E. N., Akabas, M. H. & Lummis, S. C. R. Structural and electrostatic properties of the 5-HT₃ receptor pore revealed by substituted cysteine accessibility mutagenesis. *J. Biol. Chem.* **276**, 42035–42042 (2001).
20. Panicker, S., Cruz, H., Arrabit, C. & Slesinger, P. A. Evidence for a centrally located gate in the pore of a serotonin-gated ion channel. *J. Neurosci.* **22**, 1629–1639 (2002).
21. Gunthorpe, M. J. & Lummis, S. C. Conversion of the ion selectivity of the 5-HT_{3A} receptor from cationic to anionic reveals a conserved feature of the ligand-gated ion channel superfamily. *J. Biol. Chem.* **276**, 10977–10983 (2001).
22. Unwin, N. The Croonian Lecture 2000. Nicotinic acetylcholine receptor and the structural basis of fast synaptic transmission. *Phil. Trans. R. Soc. Lond. B* **355**, 1813–1829 (2000).
23. Sokolova, O., Kolmakova-Partensky, L. & Grigorieff, N. Three-dimensional structure of a voltage-gated potassium channel at 2.5 nm resolution. *Structure* **9**, 215–220 (2001).
24. Kobertz, W. R., Williams, C. & Miller, C. Hanging gondola structure of the T1 domain in a voltage-gated K⁺ channel. *Biochemistry* **39**, 10347–10352 (2000).
25. Bass, R. B., Strop, P., Barclay, M. & Rees, D. C. Crystal structure of *Escherichia coli* MscS, a voltage-modulated and mechanosensitive channel. *Science* **298**, 1582–1587 (2002).
26. Sato, C. *et al.* The voltage-sensitive sodium channel is a bell-shaped molecule with several cavities. *Nature* **409**, 1047–1051 (2001).
27. Higgins, M. K., Weitz, D., Warner, T., Schertler, G. F. & Kaupp, U. B. Molecular architecture of a retinal cGMP-gated channel: The arrangement of the cytoplasmic domains. *EMBO J.* **21**, 2087–2094 (2002).
28. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning—A Laboratory Manual*, 2nd edn (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989).
29. Davies, P. A., Wang, W., Hales, T. G. & Kirkness, E. F. A novel class of ligand-gated ion channel is activated by Zn²⁺. *J. Biol. Chem.* **278**, 712–717 (2002).

Acknowledgements We thank M.L.J. Ashford for constructive comments on an earlier version of this manuscript. This work was supported by grants from the Wellcome Trust and Tenovus Tayside to J.A.P. and J.J.L. and from the Anonymous Trust to J.A.P.

Competing interests statement The authors declare that they have no competing financial interests

Correspondence and requests for materials should be addressed to J.A.P. (j.a.peters@dundee.ac.uk).

Viral infection switches non-plasmacytoid dendritic cells into high interferon producers

Sandra S. Diebold*, Maria Montoya†, Hermann Unger‡, Lena Alexopoulou§, Polly Roy||, Linsey E. Haswell¶, Aymen Al-Shamkhani¶, Richard Flavell§, Persephone Borrow† & Caetano Reis e Sousa*

* Immunobiology Laboratory, Cancer Research UK, London Research Institute, London WC2A 3PX, UK

† Edward Jenner Institute for Vaccine Research, Compton, Berkshire RG20 7NN, UK

‡ Institute for Virology, University of Veterinary Medicine, Vienna A-1210, Austria

§ Section of Immunobiology, Yale University School of Medicine and Howard Hughes Medical Institute, New Haven, Connecticut 06520, USA

|| London School of Hygiene & Tropical Medicine, London WC1A 7HT, UK

¶ Tenovus Research Laboratory, Cancer Sciences Division, The School of Medicine, Southampton General Hospital, Southampton SO16 6YD, UK

Type I interferons (IFN-I) are important cytokines linking innate and adaptive immunity¹. Plasmacytoid dendritic cells make high levels of IFN-I in response to viral infection and are thought to be the major source of the cytokines *in vivo*². Here, we show that conventional non-plasmacytoid dendritic cells taken from mice infected with a dendritic-cell-tropic strain of lymphocytic choriomeningitis virus make similarly high levels of IFN-I on subsequent culture. Similarly, non-plasmacytoid dendritic cells secrete high levels of IFN-I in response to double-stranded RNA (dsRNA), a major viral signature³, when the latter is introduced into the cytoplasm to mimic direct viral infection. This response is partially dependent on the cytosolic dsRNA-binding enzyme protein kinase R⁴ and does not require signalling through toll-like receptor (TLR) 3, a surface receptor for dsRNA⁵. Furthermore, we show that sequestration of dsRNA by viral NS1 (refs 6,7) explains the inability of conventional dendritic cells to produce IFN-I on infection with influenza. Our results suggest that multiple dendritic cell types, not just plasmacytoid cells, can act as specialized interferon-producing cells in certain viral infections, and reveal the existence of a TLR-independent pathway for dendritic cell activation that can be the target of viral interference.

Although all cells can produce IFN-I, plasmacytoid dendritic cells (PDCs) produce 1,000-fold higher levels than other cell types, and are responsible for systemic IFN-I responses to many viruses². However, depletion experiments show that PDCs are not essential for the IFN-I response to the Armstrong strain of lymphocytic choriomeningitis virus (LCMV)⁸, suggesting the existence of alternative interferon-producing cells *in vivo*. These could include conventional, non-plasmacytoid dendritic cells found at all potential sites of pathogen entry. To address this possibility, we purified conventional non-plasmacytoid CD11c^{high} Ly6C[−] B220[−] dendritic cells from the spleen of mice infected with LCMV intravenously, and measured IFN- α secretion after culture *in vitro*. We used two LCMV isolates, the Armstrong strain and an Armstrong variant, clone 13, which exhibit a different tropism: three days after intravenous inoculation, clone 13 can be found within dendritic cells in T-cell areas of the spleen, whereas Armstrong replicates predominantly in other cell types in the red pulp⁹. As reported previously⁸, IFN- α production was not detected on culture of conventional CD11c^{high} dendritic cells from mice infected with LCMV Armstrong (Fig. 1a). However, although little type 1 IFN messenger RNA was found in dendritic cells from clone-13-infected mice immediately on isolation (data not shown), high levels of IFN- α were produced when

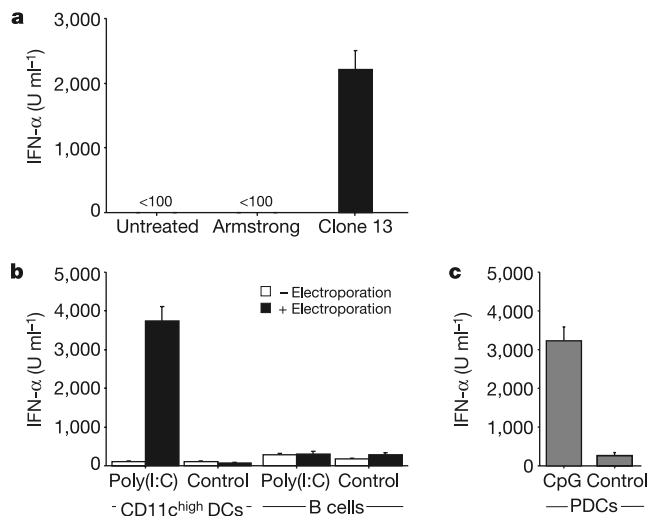


Figure 1 Viral infection or cytosolic dsRNA can induce high levels of IFN- α production by non-plasmacytoid splenic dendritic cells. **a**, C57BL/6 mice were infected intravenously with LCMV Armstrong or clone 13. Splenic CD11c^{high} dendritic cells were purified three days after infection, cultured overnight, and IFN- α accumulation in the supernatant was measured. Data are representative of three independent experiments. **b**, Splenic CD11c^{high} dendritic cells (DCs) (BALB/c) or B cells (C57BL/6) were electroporated in the presence or absence of 3 μ g poly(I:C) or were exposed to the electroporation buffer with or without poly(I:C), but not subjected to the electrical pulse, before being cultured overnight in medium alone. Data show IFN- α accumulation in supernatants. **c**, Sorted CD11c^{dim} Ly6C⁺ B220⁺ spleen PDCs (BALB/c) were cultured overnight with or without CpG (0.5 μ g ml⁻¹), and secreted IFN- α was measured as above. Data represent the average values of triplicate samples $\pm$ 1 s.d. and are representative of at least four independent experiments.

these dendritic cells were subsequently cultured *in vitro* (Fig. 1a). These results demonstrate that conventional non-plasmacytoid dendritic cells have the potential to produce high levels of IFN-I after viral exposure and, therefore, suggest that the ability to act as interferon-producing cells in viral infections may not be restricted to a single dendritic cell subtype.

The viral components leading to IFN-I induction remain largely unidentified¹⁰. Among the best known is dsRNA, an intermediate in the replicative cycle of many viruses³. To dissect some of the events linking viral recognition to IFN-I production, we examined the response of non-plasmacytoid dendritic cells to the synthetic dsRNA polyriboinosinic polyribocytidylic acid (poly(I:C)). CD11c^{high} Ly6C⁻ B220⁻ dendritic cells treated with poly(I:C) made only background levels of IFN- α (Fig. 1b). However, because viral replication takes place in the cytosol, we then assessed the effect of delivering poly(I:C) intracellularly. Electroporation with poly(I:C) allowed conventional splenic dendritic cells to produce levels of IFN- α much greater than previously reported¹¹ and in the same order of magnitude as those made by PDCs (Fig. 1b, c). In contrast, B cells (Fig. 1b) or fibroblasts (not shown) did not produce high levels of IFN- α under the same conditions. To exclude the possibility of a PDC contaminant in the CD11c^{high} spleen fraction, experiments were repeated with CD11c^{high} dendritic cells grown *in vitro* from bone marrow precursors (BM-DCs) using granulocyte macrophage colony-stimulating factor (GM-CSF). BM-DCs also made high amounts of IFN- α after electroporation with poly(I:C) or after treatment with poly(I:C) in combination with lipofectamine, a reagent for nucleic acid transfection (Fig. 2a, b). The same treatments also induced IFN- β mRNA expression (data not shown). Lipofectamine did not simply act as a nonspecific potentiator of cytokine synthesis, as interleukin-12 (IL-12) p70 induction actually decreased in the presence of the lipid (Fig. 2c). Similar to poly(I:C), dsRNA isolated directly from a reovirus also induced high levels of IFN- α production by BM-DCs when administered by electroporation or together with lipofectamine but not

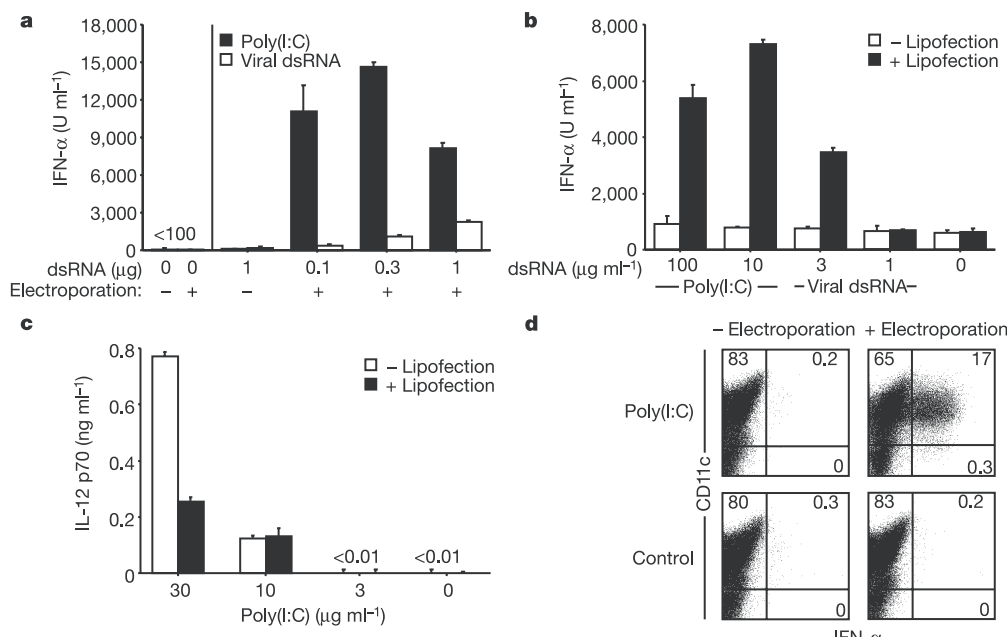


Figure 2 Cytosolic administration of dsRNA induces IFN- α production by BM-DCs. **a**, C57BL/6 BM-DCs were electroporated or not with the indicated amounts of poly(I:C) or reoviral dsRNA as in Fig. 1b. **b**, C57BL/6 BM-DCs were treated with or without different concentrations of poly(I:C) or viral dsRNA in the presence or absence of lipofectamine. **a, b**, IFN- α in culture supernatants was measured by ELISA after overnight culture. **c**, As

for **b** but measuring IL-12 p70. Data in **a-c** represent the average values of triplicate samples $\pm$ 1 s.d. and are representative of at least three independent experiments. **d**, Intracellular IFN- α staining of C57BL/6 BM-DCs after electroporation with 0.3 μ g poly(I:C). Numbers indicate percentage of total cells in each quadrant. Data are representative of three independent experiments.

when given exogenously (Fig. 2a, b; the lower levels of IFN- α induced by reoviral dsRNA are probably due to the fact that it could not be used at saturation because of limited availability). Notably, IFN- α production in BM-DC cultures was not attributable to a small number of contaminating PDCs, as 14–22% of CD11c^{high} cells stained for the cytokine after electroporation with poly(I:C) (Fig. 2d). IFN- α production by BM-DCs in response to cytosolic dsRNA could be amplified but not initiated by providing a co-signal via CD40 (Supplementary Fig. 1), suggesting that it might be further regulated during dendritic cell–T-cell interactions, as reported for other cytokines¹². We conclude that non-plasmacytoid dendritic cells can act as interferon-producing cells in response to dsRNA if the viral pattern crosses the plasma membrane.

TLR3 can mediate innate stimulation by dsRNA, and it is important for the induction of IFN- α and IFN- β genes in macrophages and for IL-12 p40 production by BM-DCs in response to poly(I:C)⁵. Notably, TLR3 message was undetectable in resting immature BM-DCs (Supplementary Fig. 2a, b). Nevertheless, activation of BM-DCs with a variety of stimuli, including poly(I:C), caused an increase in TLR3 mRNA expression (Supplementary Fig. 2b). To determine whether this might account for the IFN-I response, TLR3^{-/-} BM-DCs were treated with poly(I:C) plus lipofectamine and assayed for IFN- α secretion. As shown in Fig. 3a, TLR3 deficiency did not impair IFN- α production. To examine the involvement of other TLRs, we tested the requirement for MyD88, a TLR adapter¹³. MyD88^{-/-} BM-DCs mounted normal IFN- α responses to poly(I:C) plus lipofectamine (Fig. 3b). IFN- α induction was also resistant to the effects of a peptide originally believed to interfere specifically with TIRAP function but now known to block other pathways, including MyD88-independent TLR4 signalling, which can lead to IFN-I gene expression¹⁴ (Fig. 3b). These results, together with the known topology of TLRs (ligand-binding domain facing away from the cytosol), suggest that IFN- α

produced by dendritic cells in response to cytosolic dsRNA is TLR-independent.

The apparent lack of TLR involvement and requirement for cytosolic delivery suggested that IFN- α production might require recognition of dsRNA by an intracellular pattern recognition receptor. Protein kinase R (PKR) is a cytosolic serine/threonine kinase activated by autophosphorylation on binding to dsRNA⁴. Although PKR is primarily involved in restricting viral replication⁴, it has also been implicated in IFN responses to some viruses¹⁵ and in IFN-I induction by dsRNA in cell lines^{16–19}. Consistent with this role, 2-aminopurine, a PKR inhibitor^{16,17}, blocked IFN- α production by BM-DCs after treatment with poly(I:C) plus lipofectamine, even when CD40L was used to provide maximum amplification (Supplementary Fig. 3). The effect was selective as 2-aminopurine did not inhibit production of IL-12 p70 in response to CpG-containing oligonucleotide (CpG) or lipopolysaccharide (LPS) (see Supplementary Fig. 3). To formally implicate PKR, the experiments were repeated using BM-DCs grown from mice lacking the PKR dsRNA-binding domain (PKR^{-/-} mice¹⁸). PKR^{-/-} BM-DCs displayed a marked (although not absolute) reduction in IFN- α production after stimulation with poly(I:C) plus lipofectamine and CD40L (Fig. 3c). In contrast, IL-12 p70 production was actually increased in PKR^{-/-} dendritic cells (Fig. 3c), possibly because negative regulation by IFN-I was abrogated²⁰. We conclude that PKR has an important function in the IFN-I response of dendritic cells to cytosolic dsRNA but is not required for production of IL-12.

Many viruses have evolved strategies to avoid activation of PKR¹⁰. For example, influenza, widely used to stimulate production of IFN-I by PDCs^{21–25}, encodes the NS1 protein, which sequesters dsRNA⁷. To test the hypothesis that NS1 blocks the IFN-I response of non-plasmacytoid dendritic cells, we compared IFN- α production by BM-DCs in response to infection with wild-type

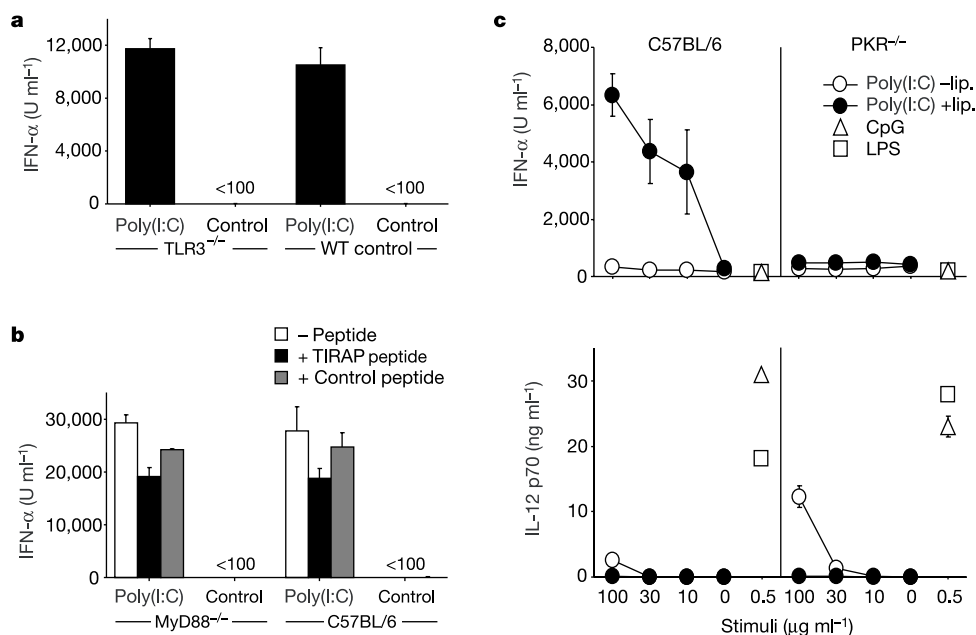


Figure 3 IFN- α induction by dsRNA does not require TLR signalling but is largely dependent on PKR. **a**, BM-DCs from TLR3^{-/-} or control mice were treated with poly(I:C) (30 μ g ml⁻¹) plus lipofectamine or left untreated (control). Cells were cultured on CD40L-expressing fibroblasts and IFN- α accumulation in culture supernatants was determined after overnight culture. **b**, MyD88^{-/-} or control wild-type (WT) BM-DCs were treated as in **a** in the presence of an inhibitory peptide (TIRAP), control peptide or in the absence of peptide. Data represent the average values of triplicate samples $\pm$ 1 s.d. and are representative of three independent experiments. The inhibitory peptide but not the

control peptide prevented upregulation of CD86 in dendritic cells treated with LPS (data not shown; see also ref. 12). **c**, BM-DCs from PKR^{-/-} or wild-type littermate control mice were treated with or without the indicated concentrations of poly(I:C) in the presence or absence of lipofectamine, or with CpG or LPS (both at 0.5 μ g ml⁻¹) in the absence of lipofectamine. Fc-CD40L was used throughout to amplify cytokine production. IFN- α and IL-12 p70 accumulation in supernatants was determined after overnight culture. Data represent the average values of triplicate samples $\pm$ 1 s.d. and are representative of four independent experiments.

influenza or with an NS1-deleted mutant virus (Δ NS1)⁶. Consistent with previous reports^{21–25}, non-plasmacytoid BM-DCs produced little IFN- α in response to wild-type influenza despite the fact that the virus infected up to 60% of the cells as assessed by staining for influenza nucleoprotein (NP) (Fig. 4a, b). However, infected BM-DCs also expressed high levels of NS1 (Fig. 4a). In contrast, infection with Δ NS1 influenza induced production of high levels of IFN- α (Fig. 4b). This was especially striking as Δ NS1 replicates poorly in wild-type cells^{6,7} and infected only a small fraction of BM-DCs (<10% NP⁺ cells) (Fig. 4a). We conclude that dsRNA produced during active viral replication in conventional dendritic cells can trigger high levels of IFN- α production provided that the virus does not actively interfere with dsRNA recognition pathways.

Our data suggest that non-plasmacytoid dendritic cells can function as interferon-producing cells after viral infection, and demonstrate that there are at least two pathways for their activation by viral dsRNA. Cell surface recognition through a PKR-independent pathway is sufficient to trigger production of IL-12 (Fig. 3c). In contrast, IFN-I production requires access of dsRNA to the cytoplasm and involves PKR and, possibly, related kinases. These results

imply that conventional dendritic cells can discriminate between direct infection with actively replicating virus compared with contact with free viral particles or virally infected cells. Notably, previous investigations have failed to reveal non-plasmacytoid dendritic cells as major IFN-I producers in response to viral infection². In some instances, this may have been due to lack of viral tropism for dendritic cells, as suggested by our experiments comparing LCMV Armstrong and clone 13 (Fig. 1a). In other cases, it is likely to have been due to lack of viral replication and consequent lack of dsRNA production (for example, experiments using inactivated viruses^{26,27}) or to viral sequestration of the stimulus as shown by our experiments with Δ NS1 influenza (Fig. 4). In contrast to conventional dendritic cells, PDCs respond both to wild-type influenza and inactivated virus^{21–25,27} and must, therefore, possess additional, dsRNA-independent pathways for IFN-I induction. Thus, the fundamental difference between plasmacytoid and non-plasmacytoid dendritic cells seems to be in the mechanisms used for coupling viral recognition to IFN-I synthesis rather than in an ontogenetically determined ability to produce vast amounts of the cytokines.

PKR has been implicated previously in IFN-I production by fibroblasts in response to exogenously added dsRNA^{16–19}. However, because TLR signalling can lead to PKR activation¹⁴, this could have been due to TLR3 engagement at the cell surface¹⁰. Here, we demonstrate that PKR can participate in induction of high levels of IFN-I independently of TLR3 or other TLRs. This may involve PKR itself acting as a cytosolic pattern recognition receptor for dsRNA or might require an upstream dsRNA-binding PKR activator such as PACT⁴. PKR can then lead to the activation of a pathway leading to phosphorylation of IRF3 and IRF7 (refs 4,19), two transcription factors critical for the expression of IFN- β /IFN- α_4 and non-IFN- α_4 genes, respectively²⁸. In contrast to fibroblasts²⁸, BM-DCs, splenic CD11c^{high} dendritic cells and splenic PDCs constitutively express mRNA for IRF3 and IRF7 (data not shown). Therefore, it is likely that most dendritic cell types can rapidly make all IFN- α subtypes on stimulation and not just the IRF3-dependent IFN- β /IFN- α_4 .

Other pathways besides PKR probably have a role in IFN-I induction in infected dendritic cells. Indeed, PKR^{−/−} dendritic cells still produce IFN- α after transfection with high levels of dsRNA (data not shown), suggesting that PKR deficiency decreases sensitivity to the stimulus (Fig. 3) but does not abrogate the response entirely. Unfortunately, this has prevented us from assessing the degree to which the IFN-I response to Δ NS1 influenza is PKR-dependent. Comparing wild-type and PKR^{−/−} dendritic cells is uninterpretable because Δ NS1 virus replicates to 10⁵ higher titres in PKR-deficient cells⁷. This generates high levels of dsRNA, which can overwhelmingly induce IFN- α production through PKR-independent pathways even if these pathways normally account for only a fraction of the IFN-I response. Such induction of IFN-I is not seen during infection with wild-type influenza presumably because dsRNA sequestration by NS1 prevents activation of all cytosolic dsRNA recognition pathways. Similar considerations may explain the data reported by ref. 19, that PKR^{−/−} fibroblasts show impaired IRF3/7 phosphorylation and IFN-I induction in response to dsRNA but not to infection with Newcastle disease virus (NDV). Although this discrepancy was originally attributed to IFN-I induction by NDV components other than dsRNA¹⁹, it may equally have been the result of unchecked viral replication in PKR-deficient cells generating levels of dsRNA sufficient to activate secondary pathways for IFN-I induction. Consistent with this interpretation, NDV-induced IRF3/7 phosphorylation was completely blocked by E3L, a vaccinia dsRNA-sequestering protein similar to NS1 (ref. 19). These arguments justify our preference for delivering controlled amounts of dsRNA rather than using virus infection to dissect IFN-I production pathways in dendritic cells. Independent of its role during viral infection, the demonstration of a TLR-independent mechanism for

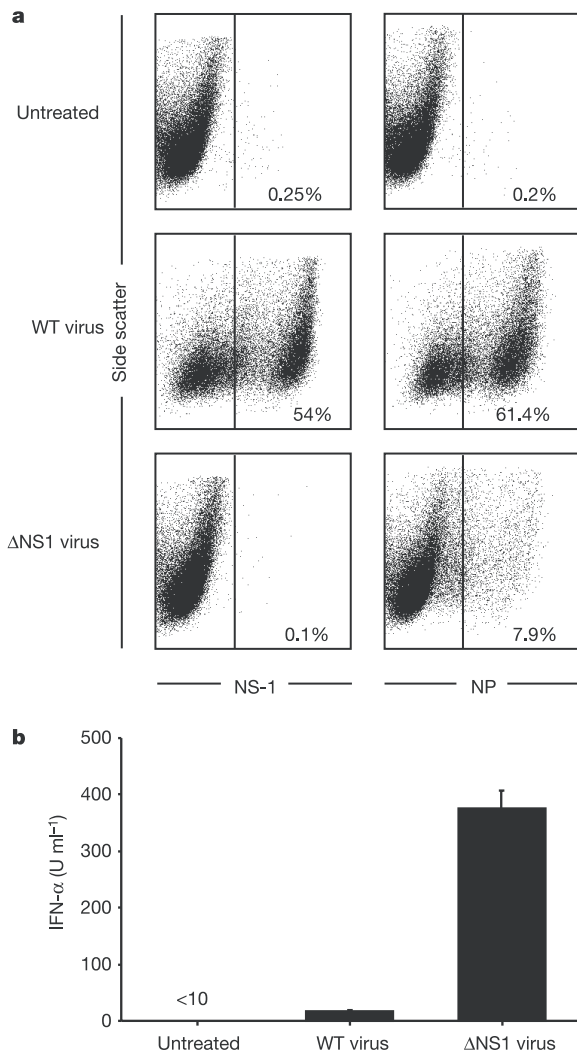


Figure 4 Infection with Δ NS1 influenza allows IFN- α production by non-plasmacytoid dendritic cells. **a**, BM-DCs were infected with influenza Δ NS1 or wild-type virus or were left untreated. Cells were stained for influenza NS1 and NP protein after overnight culture and analysed by flow cytometry after gating on CD11c^{high} cells. **b**, Same cells as in **a**. IFN- α accumulation in the supernatant was measured after overnight culture. Data are representative of three independent experiments.

cytosolic dsRNA recognition via PKR, leading to production of high amounts of IFN-I by conventional dendritic cells, suggests that cytosolic targeting of nucleic acids could be useful in vaccination and/or immunotherapy. □

Methods

Reagents

Poly(I:C) was from Pharmacia. LPS (*Escherichia coli* serotype 0128:B12) and 2-aminopurine were from Sigma. Recombinant murine GM-CSF, CpG-containing oligonucleotide and the TIRAP inhibitory and control peptides¹⁴ were made at Cancer Research UK. Reoviral dsRNA was purified from cells infected with bluetongue virus serotype 10. Soluble trimeric CD40L contained the rat CD4 leader, the α -helical coiled-coil domain of human lung surfactant protein D (amino acids 223–257) and the extracellular domain of murine CD154 (amino acids 50–260). Fc-CD40L contained the extracellular domain of murine CD154 and a modified human immunoglobulin- γ 1 (IgG1) crystallizable fragment (Fc) region. All reagents were free of endotoxin.

Cells

C57BL/6 mice were obtained from Charles River or from the Institute for Animal Health. BALB/c and MyD88^{-/-} mice¹³ (a gift from S. Akira) were bred at Cancer Research UK. PKR^{-/-} mice¹⁸ were bred at the University of Veterinary Medicine, Vienna, Austria. TLR3^{-/-} mice⁵ were bred at Yale University. Splenocytes were prepared by digestion with Liberase CI and DNaseI¹². PDC and conventional CD11c^{high} dendritic cells were purified from splenocyte suspensions as described²⁹. Sorted dendritic cells were 95–99% pure. B cells were purified by negative selection of contaminating leukocytes and were 90–95% pure. BM-DCs were generated in RPMI 1640 medium containing 10% FCS, 2 mM glutamine, 50 μ M 2-mercaptoethanol, 100 units ml⁻¹ penicillin, 100 μ g ml⁻¹ streptomycin and GM-CSF (around 1 μ g ml⁻¹). CD40L-expressing and control NIH 3T3 fibroblast cell lines were a gift from P. Hwu.

Cytokine induction

For transfection of viral dsRNA or poly(I:C), cells were seeded in triplicate in 96-well plates at a density of 0.5–2.5 $\times$ 10⁵ cells per well in 75 μ l RPMI 1640 medium without FCS or other additives; where indicated, cells were seeded onto a subconfluent monolayer of CD40L-expressing or control fibroblasts or were cultured with soluble CD40L (Fc-CD40L or CD40L trimer) to amplify the cytokine response¹². Twenty-five microlitres of transfection mix consisting of dsRNA with or without lipofectamine (GIBCO) was added to each well, and cells were incubated for 4 h. Then 100 μ l complete medium was added per well and cells were cultured overnight. For electroporation, 4 $\times$ 10⁶ cells were re-suspended in 200 μ l ‘intracellular’ buffer³⁰ with or without dsRNA and were pulsed once at 300 V, 150 μ F (Bio-Rad Gene Pulser II). Cell suspensions were then diluted with complete medium and seeded at a concentration of 2 $\times$ 10⁵ cells per well in 96-well plates with or without soluble CD40L or CD40L-expressing/control fibroblasts and were cultured overnight. Supernatant was collected at 18–20 h and assayed for IL-12 p70 and IFN- α by standard sandwich enzyme-linked immunosorbent assay (ELISA) as described^{12,27}.

Virus infection

Influenza A/PR/8/34 virus and Δ NS1 virus⁶ were used for *in vitro* infection of C57BL/6 BM-DCs. 1 $\times$ 10⁶ BM-DCs per ml were infected with virus at a multiplicity of infection of 0.7, and cell supernatants and cells were collected after 8–20 h culture. IFN- α accumulation in the supernatant was measured by ELISA as described above. Infections with LCMV were performed with Armstrong 53b strain (Armstrong) or clone 13. Mice were infected intravenously with 2 $\times$ 10⁶ plaque-forming units per mouse in 200 μ l volume. Splenic dendritic cells were purified 3 days after infection and cultured *in vitro* at 2 $\times$ 10⁶ cells per ml in 200 μ l complete medium with 20 ng ml⁻¹ GM-CSF. IFN- α accumulation in the supernatant after 18–20 h culture was measured using an IFN- α ELISA kit from PBL (Alexis Corporation).

Flow cytometry

For intracellular IFN- α staining, dendritic cells were electroporated with poly(I:C) and cultured for 3 h. Brefeldin A (Sigma; 5 μ g ml⁻¹) was added for an additional 3 h and cells were collected and fixed in paraformaldehyde. Staining was performed in saponin-containing buffer using a mixture of rat anti-mouse IFN- α antibodies (clone F18, Hycult Biotechnology b.v. and clone RMMA-1, PBL Biomedical Laboratories), followed by biotinylated mouse anti-rat IgG (Jackson ImmunoResearch) and Streptavidin-PE (PharMingen). Cells were subsequently stained for CD11c (clone HL3, PharMingen). For intracellular staining of influenza NP and NS1 proteins, infected dendritic cells were fixed and stained as above using mouse anti-NP antibody or mouse anti-NS1 antibody followed by biotinylated rat anti-mouse IgG (Jackson ImmunoResearch) and Streptavidin-PE. Cells were subsequently stained for CD11c.

PCR

BM-DCs or CD8 α ⁺ dendritic cells were analysed for TLR3 mRNA expression as described²⁹.

Received 19 March; accepted 28 May 2003; doi:10.1038/nature01783.
Published online 22 June 2003.

1. Le Bon, A. & Tough, D. F. Links between innate and adaptive immunity via type I interferon. *Curr. Opin. Immunol.* **14**, 432–436 (2002).

- Colonna, M., Krug, A. & Cella, M. Interferon-producing cells: on the front line in immune responses against pathogens. *Curr. Opin. Immunol.* **14**, 373–379 (2002).
- Jacobs, B. L. & Langland, J. O. When two strands are better than one: the mediators and modulators of the cellular responses to double-stranded RNA. *Virology* **219**, 339–349 (1996).
- Williams, B. R. Signal integration via PKR. *Sci. STKE* **2001**(89), RE2 (2001); at (http://stke.sciencemag.org/cgi/content/full/OC_sigtrans;2001/89/re2).
- Alexopoulou, L., Holt, A. C., Medzhitov, R. & Flavell, R. A. Recognition of double-stranded RNA and activation of NF- κ B by Toll-like receptor 3. *Nature* **413**, 732–738 (2001).
- Garcia-Sastre, A. *et al.* Influenza A virus lacking the NS1 gene replicates in interferon-deficient systems. *Virology* **252**, 324–330 (1998).
- Bergmann, M. *et al.* Influenza virus NS1 protein counteracts PKR-mediated inhibition of replication. *J. Virol.* **74**, 6203–6206 (2000).
- Dalod, M. *et al.* Interferon alpha/beta and interleukin 12 responses to viral infections: pathways regulating dendritic cell cytokine expression *in vivo*. *J. Exp. Med.* **195**, 517–528 (2002).
- Borrow, P., Evans, C. F. & Oldstone, M. B. Virus-induced immunosuppression: immune system-mediated destruction of virus-infected dendritic cells results in generalized immune suppression. *J. Virol.* **69**, 1059–1070 (1995).
- Katze, M. G., He, Y. & Gale, M. Viruses and interferon: A fight for supremacy. *Nature Rev. Immunol.* **2**, 675–687 (2002).
- Hochrein, H. *et al.* Differential production of IL-12, IFN- α , and IFN- γ by mouse dendritic cell subsets. *J. Immunol.* **166**, 5448–5455 (2001).
- Edwards, A. D. *et al.* Microbial recognition via toll-like receptor-dependent and -independent pathways determines the cytokine response of murine dendritic cell subsets to CD40 triggering. *J. Immunol.* **169**, 3652–3660 (2002).
- Adachi, O. *et al.* Targeted disruption of the MyD88 gene results in loss of IL-1- and IL-18-mediated functions. *Immunity* **9**, 143–150 (1998).
- Horng, T., Barton, G. M. & Medzhitov, R. TIRAP: an adapter molecule in the Toll signaling pathway. *Nature Immunol.* **2**, 835–841 (2001).
- Balachandran, S. *et al.* Essential role for the dsRNA-dependent protein kinase PKR in innate immunity to viral infection. *Immunity* **13**, 129–141 (2000).
- Marcus, P. I. & Sekellick, M. J. Interferon induction by viruses. XVI. 2-Aminopurine blocks selectively and reversibly an early stage in interferon induction. *J. Gen. Virol.* **69**, 1637–1645 (1988).
- Zinn, K., Keller, A., Whittemore, L. A. & Maniatis, T. 2-Aminopurine selectively inhibits the induction of β -interferon, c-fos, and c-myc gene expression. *Science* **240**, 210–213 (1988).
- Yang, Y. L. *et al.* Deficient signaling in mice devoid of double-stranded RNA-dependent protein kinase. *EMBO J.* **14**, 6095–6106 (1995).
- Smith, E. J., Marie, I., Prakash, A., Garcia-Sastre, A. & Levy, D. E. IRF3 and IRF7 phosphorylation in virus-infected cells does not require double-stranded RNA-dependent protein kinase R or I κ B kinase but is blocked by Vaccinia virus E3L protein. *J. Biol. Chem.* **276**, 8951–8957 (2001).
- Cousens, L. P., Orange, J. S., Su, H. C. & Biron, C. A. Interferon- α/β inhibition of interleukin 12 and interferon- γ production *in vitro* and endogenously during viral infection. *Proc. Natl Acad. Sci. USA* **94**, 634–639 (1997).
- Cella, M. *et al.* Plasmacytoid monocytes migrate to inflamed lymph nodes and produce large amounts of type I interferon. *Nature Med.* **5**, 919–923 (1999).
- Kadowaki, N., Antonenko, S., Lau, J. Y. & Liu, Y. J. Natural interferon alpha/beta-producing cells link innate and adaptive immunity. *J. Exp. Med.* **192**, 219–226 (2000).
- Nakano, H., Yanagita, M. & Gunn, M. D. Cd11c(+)b220(+)gr-1(+) cells in mouse lymph nodes and spleen display characteristics of plasmacytoid dendritic cells. *J. Exp. Med.* **194**, 1171–1178 (2001).
- Bruno, L., Seidl, T. & Lanzavecchia, A. Mouse pre-immunocytes as non-proliferating multipotent precursors of macrophages, interferon-producing cells, CD8 α (+) and CD8 α (-) dendritic cells. *Eur. J. Immunol.* **31**, 3403–3412 (2001).
- O’Keeffe, M. *et al.* Mouse plasmacytoid cells: long-lived cells, heterogeneous in surface phenotype and function, that differentiate into CD8(+) dendritic cells only after microbial stimulus. *J. Exp. Med.* **196**, 1307–1319 (2002).
- Siegal, F. P. *et al.* The nature of the principal type 1 interferon-producing cells in human blood. *Science* **284**, 1835–1837 (1999).
- Asselin-Paturel, C. *et al.* Mouse type I IFN-producing cells are immature APCs with plasmacytoid morphology. *Nature Immunol.* **2**, 1144–1150 (2001).
- Marie, I., Durbin, J. E. & Levy, D. E. Differential viral induction of distinct interferon- α genes by positive feedback through interferon regulatory factor-7. *EMBO J.* **17**, 6660–6669 (1998).
- Edwards, A. D. *et al.* Toll-like receptor expression in murine DC subsets: lack of TLR7 expression by CD8 α ⁺ DC correlates with unresponsiveness to imidazoquinolines. *Eur. J. Immunol.* **33**, 827–833 (2003).
- van den Hoff, M. J., Christoffels, V. M., Labruyere, W. T., Moorman, A. F. & Lamers, W. H. Electrotransfection with “intracellular” buffer. *Methods Mol. Biol.* **48**, 185–197 (1995).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This work was funded by Cancer Research UK. We thank T. Muster for the Δ NS1 virus; K. Thielemans for suggestions on the electroporation protocol; M. Edwards for assistance with the LCMV experiments; M. Albert for advice on the influenza experiments; C. Asselin-Paturel for details of the IFN- α ELISA; and B. Strobl for advice on PKR^{-/-} mice. We are grateful to R. Germain, A. O’Garra, G. Schiavo, D. Cantrell, F. Batista and members of the Immunobiology Laboratory, Cancer Research UK, for advice and critical review of the manuscript. S.S.D. is supported by an EMBO fellowship and a Cancer Research UK postdoctoral fellowship. This is publication number 64 from The Edward Jenner Institute for Vaccine Research. R.F. is an investigator of the Howard Hughes Medical Institute.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.R.S. (caetano@cancer.org.uk).

Impairment of dendritic cells and adaptive immunity by anthrax lethal toxin

Anshu Agrawal*, Jai Lingappa†, Stephen H. Leppla‡, Sudhanshu Agrawal*, Abdul Jabbar*, Conrad Quinn† & Bali Pulendran*

* Emory Vaccine Research Center, 954 Gatewood Road, Atlanta, Georgia 30329, USA

† Meningitis and Special Pathogens Branch, Division of Bacterial and Mycotic Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia 30333, USA

‡ Microbial Pathogenesis Section, NIAID, Bethesda, Maryland 20892-4350, USA

Anthrax poses a clear and present danger as an agent of biological terrorism^{1–3}. Infection with *Bacillus anthracis*, the causative agent of anthrax, if untreated can result in rampant bacteraemia, multisystem dysfunction and death^{4–8}. Anthrax lethal toxin (LT) is a critical virulence factor of *B. anthracis*, which occurs as a complex of protective antigen and lethal factor. Here we demonstrate that LT severely impairs the function of dendritic cells—which are pivotal to the establishment of immunity against pathogens—and host immune responses by disrupting the mitogen-activated protein (MAP) kinase intracellular signalling network. Dendritic cells exposed to LT and then stimulated with lipopolysaccharide do not upregulate co-stimulatory molecules, secrete greatly diminished amounts of proinflammatory cytokines, and do not effectively stimulate antigen-specific T cells *in vivo*. Furthermore, injections of LT induce a profound impairment of antigen-specific T- and B-cell immunity. These data suggest a role for LT in suppressing host immunity during *B. anthracis* infections, and represent an immune evasion strategy, where a microbe targets MAP kinases in dendritic cells to disarm the immune response.

Anthrax LT constitutes a complex of lethal factor (LF) with a heptamer of protective antigen (PA)^{1–8}. PA binds to a ubiquitous receptor on the cell surface, and facilitates entry of LF into the cell⁹. Recent studies indicate that LT disrupts the MAP kinase signalling pathway in macrophages^{10–13}. MAP kinases are an ancient signal transduction pathway in mammalian cells, and comprise three major groups: the extracellular signal-regulated protein kinases (ERK1 and ERK2), the p38 MAP kinases, and the c-Jun amino-terminal kinases (JNK)¹⁴. LT proteolytically cleaves short N-terminal fragments from mitogen or extracellular signal-regulated kinase 1 (MEK1) MEK2 and MEK3, the upstream activators of ERK1, ERK2 and p38, respectively, and inhibits macrophages^{10–13}. In contrast, other studies indicate that LT induces macrophages to secrete proinflammatory cytokines *in vivo*¹⁵. Despite this knowledge, there is no information about the effects of *B. anthracis* or its toxins on antigen presentation or the adaptive immune response. Such information is essential to understand the pathogenesis of anthrax and devise new therapeutic strategies.

Dendritic cells are the most efficient antigen-presenting cells in the immune system^{16–19}. They have critical roles in initiating and regulating the adaptive immune response, and are scattered throughout the body, including the various portals of microbe entry, where they reside in an immature form^{16–19}. Immature dendritic cells are 'immunological sensors', alert for potentially dangerous microbes, and are capable of decoding and integrating such signals. They then transport this information to naive T cells in the secondary lymphoid organs, undergoing a maturation process en route. Here, the mature dendritic cells present this information to T cells, thus launching an immune response^{16–19}. Dendritic cells can also tune the immune response by modulating the amplitude or

the type of the response^{16–19}. There is currently no knowledge about the interaction between *B. anthracis* components and dendritic cells. Given their pivotal function in immune modulation, it is imperative to understand how dendritic cell function is manipulated by *B. anthracis*, particularly because: (1) impairment of dendritic cells and adaptive immunity during the early stages of *B. anthracis* infection will permit unrestrained spread of bacteria and progression of the disease; and (2) the public health context of post-exposure susceptibility to unrelated infections, where impairment of dendritic cells might compromise immunity against such infections²⁰. Therefore, we investigated the effects of LT on murine and human dendritic cells, and the subsequent effects on the adaptive immune response.

First, we determined the impact of LT exposure on the function of dendritic cells *in vitro* (Fig. 1). Mouse spleen dendritic cells were isolated from mice injected with a plasmid DNA encoding the *in vivo* dendritic cell growth factor Flt3 ligand (FL)—in order to increase the number of dendritic cells *in vivo*^{21,22}—and were exposed to PA or PA plus various concentrations of LF, for 6 h. Cells were then washed and stimulated with lipopolysaccharide (LPS) overnight. The secretion of tumour necrosis factor (TNF)- α , interleukin (IL)-1 α , IL-6 and IL-12 was severely impaired in LT-treated dendritic cells, in a dose-dependant manner (Fig. 1a). Furthermore, the upregulation of the co-stimulatory molecules CD40, CD80 and CD86, which are essential for the induction of adaptive immunity, was severely impaired by LT exposure (Fig. 1c). Similar results were obtained using peptidoglycan, a stimulus from Gram-positive bacteria (Supplementary Fig. 1). LT can induce cell death in macrophages^{10–13} (Supplementary Fig. 2). To address whether the suppressive effects on dendritic cells were also due to LT-induced cell death, mouse spleen dendritic cells were cultured with PA, or PA plus LF, for 6 h, and the viability determined by the detection of biologically active caspase 3, a signature marker for cells committed to the apoptotic pathway²³ (Fig. 1b), and through Trypan blue staining (Supplementary Fig. 2). Notably, LT did not induce enhanced apoptosis in dendritic cells, even at 48 h (Fig. 1b; see also Supplementary Fig. 2). In contrast, dendritic cells that were not exposed to LT, but were repeatedly frozen and thawed, or exposed to ultraviolet radiation (Fig. 1b) contained a significant proportion of apoptotic cells. As described before^{10–13}, LT rapidly killed macrophages (Supplementary Fig. 2). Thus, LT seems to exert markedly different effects on dendritic cells and macrophages *in vitro*.

Given these surprising effects on dendritic cells, we addressed whether LT-exposed dendritic cells were able to prime T cells. Briefly, LT-treated dendritic cells were cultured with naive, allogeneic CD4⁺ T cells, and the proliferation of T cells was determined by thymidine labelling. As shown in Fig. 1d, LT-treated dendritic cells are impaired in their ability to prime allogeneic CD4⁺ T cells *in vitro*. Moreover, re-stimulation of these CD4⁺ T cells with a potent T-cell mitogen, phorbol 12-myristate 13-acetate (PMA) plus ionomycin, failed to induce T-helper cell subtype 1 (T_H1; interferon (IFN)- γ) or T_H2 (IL-4, IL-5 and IL-10) cytokines from these cells (Fig. 1e). Therefore, LT-treated dendritic cells are unable to induce efficient priming of naive T cells. Similar effects were observed with human monocyte-derived dendritic cells (MDDCs)^{16–19} (Supplementary Fig. 3).

Given these potent effects on dendritic cells *in vitro*, we determined whether LT-primed dendritic cells could induce T-cell unresponsiveness *in vivo*, using a transgenic mouse model in which the fate of antigen-specific CD4⁺ T cells could be tracked *in vivo*²⁴. The model involves DO11.10 mice, which express transgenic T-cell receptors (TCRs) that recognize an ovalbumin peptide (OVA) in the context of the class II major histocompatibility complex (MHC) I-A^d molecule²⁵. These antigen-specific T cells can be detected using an antibody that is specific to this TCR^{24,26}. CD4⁺ T cells from these mice were injected into syngeneic BALB/c mice, such that the donor cells constituted a small but detectable

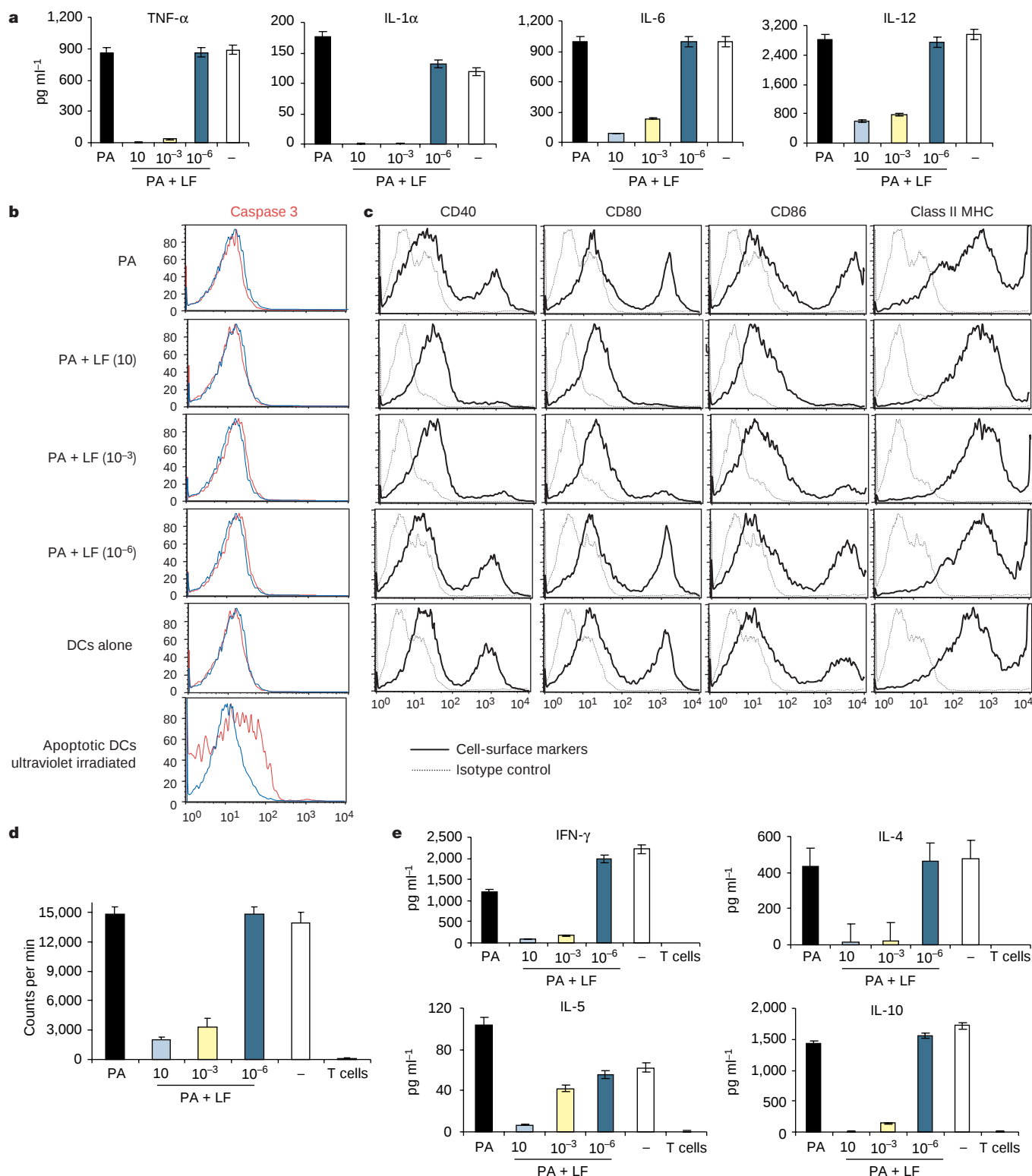


Figure 1 LT impairs dendritic cell function *in vitro*. **a**, Spleen dendritic cells (DCs) were isolated and cultured for 6 h with either PA alone (200 ng ml⁻¹), or PA (200 ng ml⁻¹) plus the indicated concentrations of LF (10, 10⁻³ or 10⁻⁶ μ g ml⁻¹), or with medium alone. Then the cells were washed and stimulated with LPS for 24 h. Cytokines secreted were measured by ELISA. Differences in cytokine levels between the PA plus LF (10 or 10⁻³ groups) and the PA or dendritic cell groups are highly significant ($P < 0.001$). **b, c**, The expression of biologically active caspase 3 (**b**) and co-stimulatory molecules (**c**) on gated CD11c⁺ dendritic cells was determined by flow cytometry. **d, e**, Spleen

dendritic cells were cultured for 6 h with the indicated conditions, then washed and cultured with allogeneic CD4⁺ T cells. T-cell proliferation was determined by [³H] labelling (**d**) and T-cell cytokine secretion was determined after an overnight re-stimulation with PMA plus ionomycin (**e**). Differences in T-cell proliferation and IFN- γ , IL-4 and IL-10 levels between the PA plus LF (10 or 10⁻³) groups and the other PA or dendritic cell groups are highly significant ($P < 0.001$). The results are representative of ten independent experiments.

fraction in the host. The reconstituted mice can then be immunized with OVA plus an adjuvant to study the dynamics of CD4⁺ T-cell responses *in vivo*^{24,26}. Spleen dendritic cells were isolated, exposed to PA or PA plus LF for 4 h in the presence or absence of OVA, washed, and then injected into the footpads of BALB/c mice reconstituted with DO11.10 cells. Four days later, at the height of the immune response^{24,26}, the clonal expansion of OVA-specific CD4⁺ T cells in the draining lymph nodes was determined. Dendritic cells pulsed with OVA alone, or with PA plus OVA, induce significant clonal expansion of KJ126⁺ CD4⁺ T cells, but dendritic cells exposed to PA plus LF plus OVA are impaired in their ability to prime such cells (Fig. 2a).

Next, we determined whether OVA-specific CD4⁺ T cells, which had been primed *in vivo*, could be re-stimulated *in vitro*. Thus, we cultured single cell suspensions from the draining lymph nodes of the various groups with different concentrations of OVA. Whereas *in vivo* stimulation with dendritic cells plus OVA or dendritic cells plus PA plus OVA results in vigorous proliferation *in vitro*, *in vivo* stimulation with dendritic cells plus PA plus LF plus OVA does not (Fig. 3b). Furthermore, consistent with the *in vitro* findings reported above (Fig. 1e), *in vivo* stimulation with dendritic cells plus PA plus LF plus OVA results in a profound diminution of T_H1 and T_H2 cytokines *in vivo* (Fig. 3c). Therefore, LT-treated dendritic

cells fail to prime antigen-specific CD4⁺ T cells *in vivo*. Notably, dendritic cells plus PA plus OVA induces a significantly weaker (Fig. 1b) but more polarized T_H1 response (Fig. 1c) than dendritic cells plus OVA. It should be noted that the response stimulated by dendritic cells plus PA plus LF plus OVA is similar to that induced by dendritic cells alone, and much lower than that induced by dendritic cells plus OVA, or dendritic cells plus PA plus OVA. This suggests that although the dendritic cells plus PA plus LF plus OVA group contains OVA, the dendritic cells are still unable to induce productive immunity. Nevertheless, KJ126⁺ CD4⁺ T cells stimulated by dendritic cells plus PA plus LF plus OVA display increased forward and side scatter, and enhanced expression of CD25, CD69 and CD44, together with downregulation of CD62-L (Supplementary Fig. 4), early signatures of T-cell activation. This suggests that such T cells are initially activated, but fail to differentiate into memory cells, reminiscent of the aborted T-cell activation induced by soluble OVA^{24,26}.

Given its inhibitory effects on dendritic cells, we investigated the effect of LT on adaptive immune responses *in vivo* (Fig. 4). Cohorts of BALB/c mice were injected with PBS, PA alone or PA plus LF, intraperitoneally (i.p.), and then challenged i.p. with ovalbumin protein adsorbed on alum, 6 h later. Four days later, the spleens were collected, and single cell suspensions cultured with graded concen-

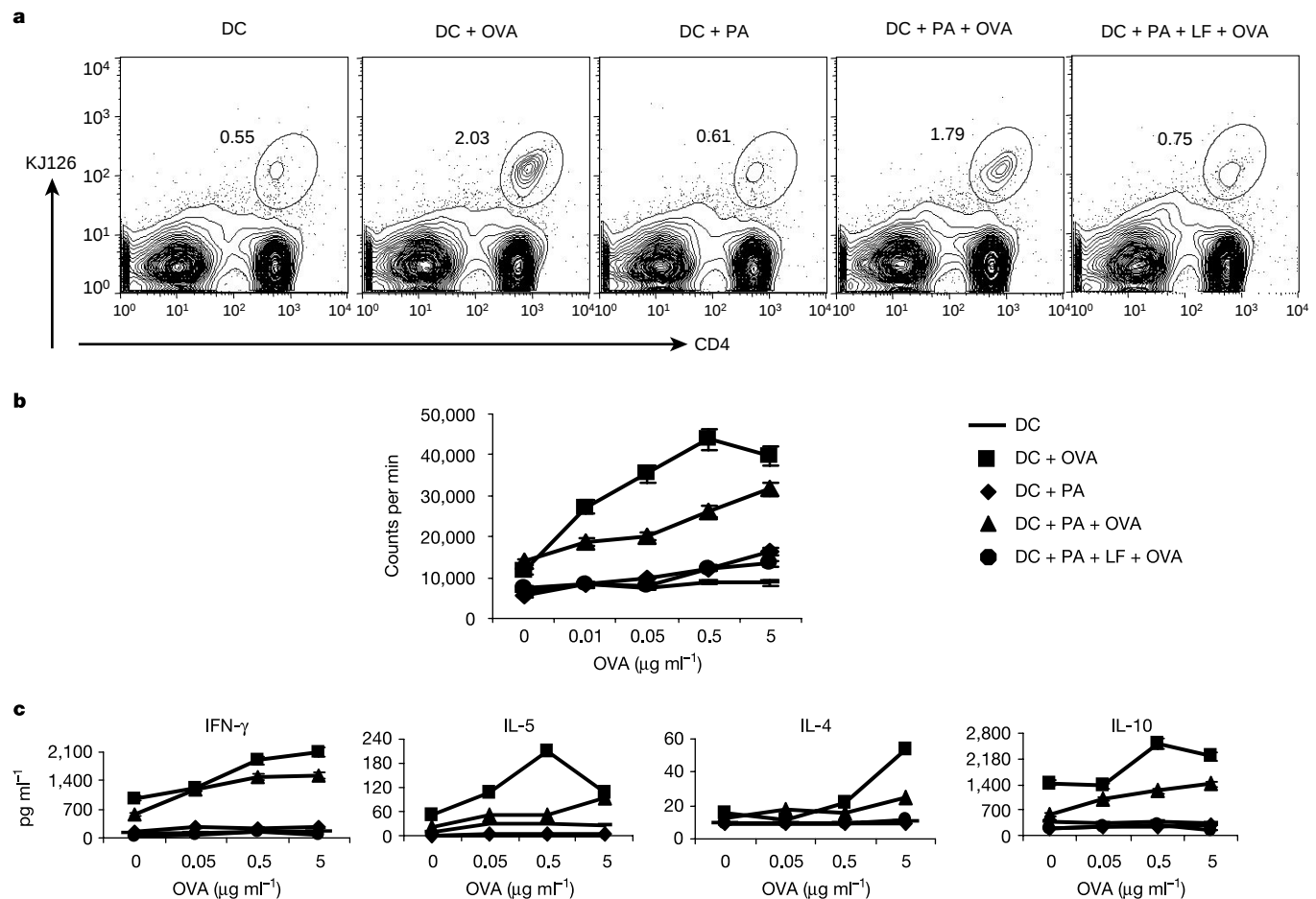


Figure 2 LT-treated dendritic cells impair T-cell responses *in vivo*. Spleen dendritic cells were cultured in the presence of either PA alone (200 ng ml⁻¹), or PA (200 ng ml⁻¹) plus LF (10 $\mu\text{g ml}^{-1}$) for 6 h, washed, loaded with OVA(323–339) peptide and injected into the footpads of BALB/c mice reconstituted with DO11.10 cells. **a**, After 4 days, clonal expansion of OVA-specific CD4⁺ T cells (KJ126⁺ CD4⁺) in the draining lymph nodes was determined by flow cytometry. **b**, **c**, Lymph node cells from the different cohorts of mice

were cultured with the indicated concentrations of OVA(323–339) for 72 h, and then T-cell proliferation was determined by [³H] labelling (**b**) and cytokine secretion (**c**) measured by ELISA. Differences in T-cell proliferation (Fig. 3b) and cytokine levels (Fig. 3c) between the DC + OVA group and the DC + PA + LF + OVA group is highly significant ($P < 0.01$). The results are representative of five independent experiments.

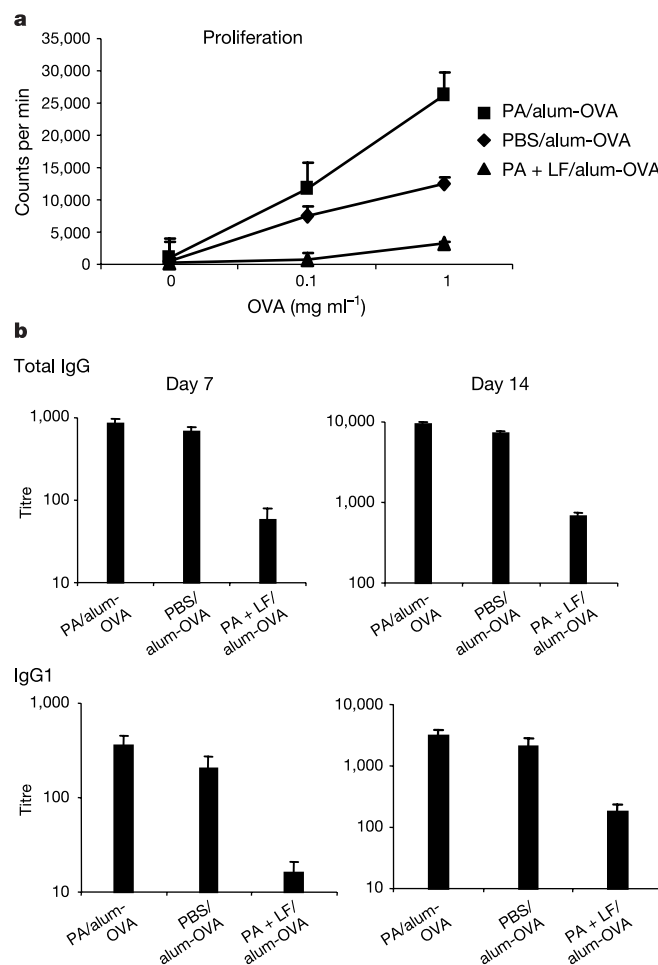


Figure 3 LT impairs antigen-specific T- and B-cell responses *in vivo*. BALB/c mice were injected i.p. with either PBS, PA alone (100 µg), or PA (100 µg) plus LF (7.5 µg), and then challenged i.p. with 500 µg ovalbumin protein adsorbed on alum, 6 h later. **a**, Four days later, the spleens were collected and single cell suspensions cultured in the presence of graded concentrations of ovalbumin. The *in vitro* proliferation of the ovalbumin-specific T cells was determined by thymidine labelling, 72 h later. The

difference between the PBS and PA alone groups versus the PA plus LF group is highly significant ($P < 0.01$). **b**, The ovalbumin-specific IgG and IgG1 antibody titres, at 7 and 14 days after immunization with ovalbumin plus alum, was determined. The difference between the PBS and PA alone groups versus the PA plus LF group is highly significant ($P < 0.001$). The results are representative of three independent experiments, each with six mice per group.

trations of ovalbumin. The *in vitro* proliferation of ovalbumin-specific T cells were determined. Notably, PA alone enhanced the response significantly. However, there is a profound impairment of ovalbumin-specific proliferation by pre-injection of PA plus LF (Fig. 4a). In order to determine whether this impairment of T-cell response resulted in antigen-specific B-cell responses, we determined the ovalbumin-specific antibody titres at 7 and 14 days after immunization with ovalbumin plus alum. As shown in Fig. 4b, there is a highly significant reduction in ovalbumin-specific immunoglobulin-γ (IgG) and IgG1 antibody titres in mice pre-injected with LT. Taken together, these data suggest that LT severely impairs dendritic cell function to disarm the adaptive immune response.

What is the mechanism of action of LT on dendritic cells? LT can specifically cleave certain MEKs in macrophages, resulting in abrogated p38 and ERK signalling^{10–13}. To determine whether the suppressive effects on dendritic cells were mediated by means of such a mechanism, we assessed the phosphorylation of p38 and ERK1/2 in LT-treated dendritic cells re-stimulated with LPS. Exposure to 10 or 10⁻³ µg ml⁻¹ LF severely impairs the phosphorylation of p38 and ERK1/2 (Fig. 4a). Given these effects on the MAP kinase pathway in dendritic cells, we investigated whether we could

mimic the effects of LT using synthetic MAP kinase inhibitors. Thus we cultured mouse dendritic cells with SB203580 (a p38 kinase inhibitor)^{27,28}, PD98059 (a MEK1 inhibitor)^{27,28}, or UO126 (a MEK1 and MEK2 inhibitor)^{27,28}, or with a combination of the three inhibitors. Although any of the inhibitors alone had modest effects, a combination of the three inhibitors was very potent at inhibiting cytokine production in a manner similar to LT (Fig. 4b). Furthermore, we performed additional experiments using a mutant form of LF (LF(E687C)), which has a point mutation in its catalytic site that enables it to bind MEK1–3, but that renders it unable to catalyse cleavage of MEK1–3 (ref. 29). This mutant LF does not suppress dendritic cell function (Fig. 4c), suggesting that cleavage of MEKs is essential for the effects observed.

The present data suggest that LT, the major virulence factor of *B. anthracis*, exerts profound inhibitory effects on dendritic cells, ultimately resulting in impaired adaptive immunity. This, to the best of our knowledge, is the first evidence of any interaction between *B. anthracis* and dendritic cells, and suggests an immune evasion mechanism that disrupts the MAP kinase signalling pathway in dendritic cells to foil adaptive immunity. Such a mechanism might operate early in infection, when the bacterial burden and LT concentrations are at sublethal levels. At advanced stages of infec-

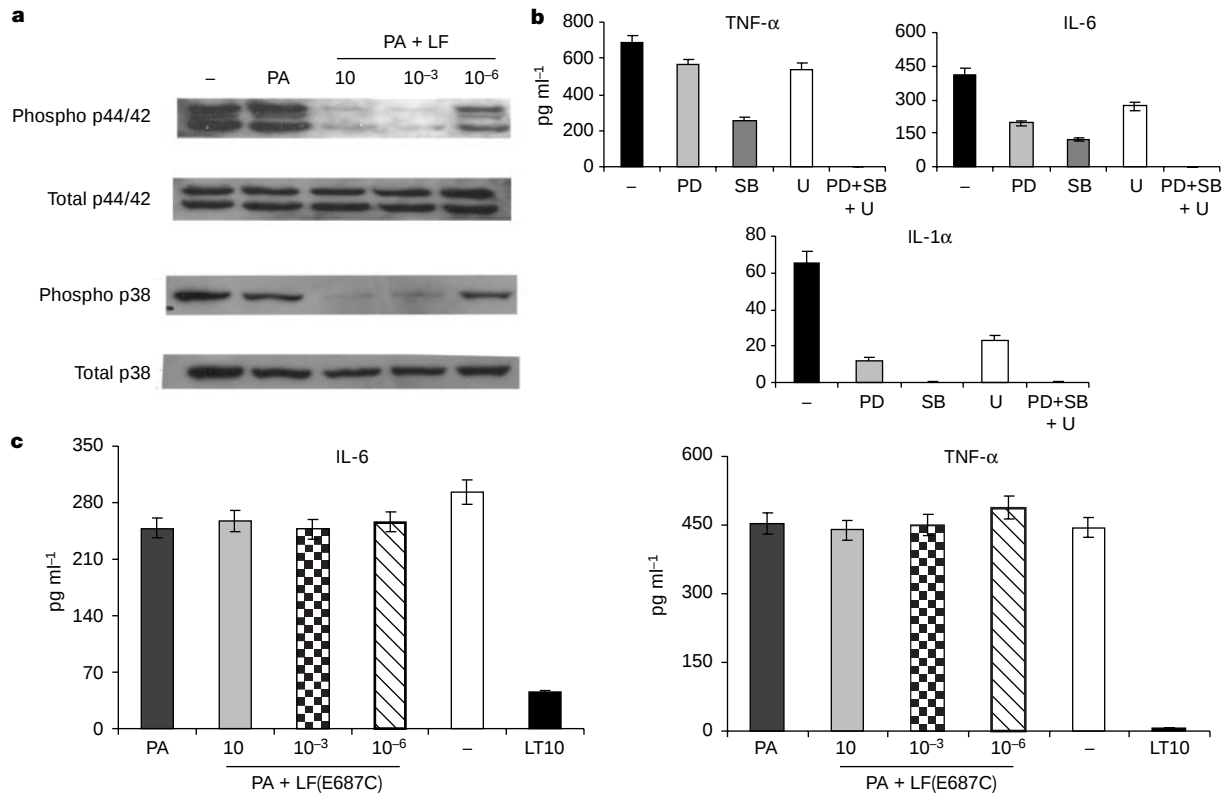


Figure 4 LT impairs phosphorylation of p38 and ERK1/2 in dendritic cells. **a**, Spleen dendritic cells were cultured in the presence of either PA alone (200 ng ml⁻¹) or PA (200 ng ml⁻¹) plus LF (10 to 10⁻⁶ μg ml⁻¹) for 6 hrs, washed, and stimulated with LPS for 30 min. The cells were lysed and then phosphorylation of p38 and ERK1/2 was determined by western blot. **b**, Inhibition of proinflammatory cytokines by synthetic MAP kinase inhibitors. Spleen dendritic cells were cultured for 1 h with the ERK1/2

inhibitors U0126 (U) or PD98059 (PD), the p38 inhibitor SB203580 (SB), or with a combination (PD + SB + U). Then the cells were stimulated for 24 h with LPS, and cytokine secretion was analysed by ELISA. **c**, LF(E687C), a catalytically inactive mutant of LF, does not impair dendritic cell function. Spleen dendritic cells were cultured with LF(E687C), as described in **a**. The results are representative of three experiments.

tion, characterized by bacteraemia and toxic shock, LT may exert quite different inflammatory effects. The mechanism of transition from a situation of impaired immunity, early in infection, to one of toxic shock-like symptoms, late in infection, is a puzzle that clearly deserves further investigation. Finally, LT or its derivatives may also represent potential therapeutic tools in the regulation of deleterious immune responses in autoimmunity or transplantation. This new therapeutic approach may be especially feasible, given the recent elucidation of the crystal structures of LF²⁹ and PA³⁰.

Methods

Mice

DO11.10 TCR transgenic mice and BALB/c mice were purchased from Jackson Laboratory.

Flt3-ligand-encoding plasmid

This plasmid was constructed as previously described²². A total of 15 μg was mixed in 3 ml saline and injected intravenously to induce dendritic cell expansion *in vivo*.

Isolation and culture of spleen dendritic cells

CD11c⁺ dendritic cells were purified from spleens as described previously²¹. Briefly, spleens of FL-injected BALB/c mice were cut into small fragments and digested with collagenase type 4 (1 mg ml⁻¹; Worthington Biochemicals) in DMEM medium supplemented with 1% fetal calf serum for 30 min at 37 °C. Cells were washed twice in PBS, 2 mM EDTA, 5% fetal calf serum. Then, dendritic cells were enriched using CD11c⁺ microbeads from Miltenyi Biotec. This enrichment process was repeated, and the resulting purity of CD11c⁺ cells after two rounds was >95%. Dendritic cells were cultured for 6 h with either PA alone (200 ng ml⁻¹), or PA (200 ng ml⁻¹) plus the indicated concentrations of LF (10, 10⁻³ or 10⁻⁶ μg ml⁻¹), or with medium alone. Then the cells were washed and stimulated with LPS for 24 h. secreted cytokines were measured by enzyme-linked immunosorbent assay (ELISA) kits (Pharmingen). The expression of

caspase 3 (reagent from OncoImmunin Inc.) and co-stimulatory molecules (antibodies from Pharmingen) on gated CD11c⁺ dendritic cells was determined by flow cytometry²³. In other experiments, spleen dendritic cells were cultured for 6 h with the various indicated conditions, then washed and cultured with allogeneic CD4⁺ T cells. T-cell proliferation was determined by [³H] labelling and T-cell cytokine secretion was determined after an overnight re-stimulation with PMA plus ionomycin.

Isolation of peritoneal macrophages

Peritoneal macrophages were collected by injecting 10 ml cold, sterile PBS into the peritoneal cavity followed by gentle massage and removal of lavage fluid. The fluid was centrifuged at low speed, and re-suspended in DMEM medium, supplemented with 10% fetal calf serum. Cells were counted, plated into tissue culture plates, and incubated for 2 h at 37 °C, 5% CO₂. Non-adherent cells were discarded, and the adherent cells (>90% macrophages) were used in experiments.

Caspase assay for apoptosis

Apoptosis was assayed by specific cleavage of cell-permeable, fluorogenic caspase substrates²³. These substrates are composed of two fluorophores covalently linked to peptides comprising 18 amino acids containing the proteolytic cleavage sites for individual caspases. In the uncleaved substrates, fluorescence is quenched owing to the formation of intramolecular excitonic dimers. On cleavage of the peptides by specific caspases, the fluorophore-fluorophore interaction is abolished, leading to an increase in fluorescence that can be detected by flow cytometry.

Generation and culture of human MDDCs

CD14⁺ monocytes were enriched from peripheral blood mononuclear cells using an enrichment step, and cultured in six-well plates (1 × 10⁶ per well) for 6 days with recombinant human granulocyte/macrophage colony-stimulating factor at 100 ng ml⁻¹ (PeproTech) plus recombinant human IL-4 at 20 ng ml⁻¹ (PeproTech). At day 6, the cultures consisted uniformly of CD1a⁺ CD14⁺, HLA-DR⁺ CD11c⁺ cells, which were negative for CD83, the human dendritic cell maturation marker, and expressed intermediate levels of co-stimulatory molecules as CD86. On day 6, the immature MDDCs were pulsed with PA (200 ng ml⁻¹), or PA (200 ng ml⁻¹) plus various concentrations of LF, for 6 h. Then the cells were washed and either re-stimulated with LPS for 24 h (for

cytokines), or washed and cultured at graded doses with 1×10^5 naive CD4⁺ CD45RA⁺; proliferation assessed by [³H] labelling. Cytokine secretion by the T cells was determined after overnight re-stimulation with PMA plus ionomycin.

Analyses of *in vivo* immune response

For adoptive transfers, age-matched BALB/c recipients were reconstituted with 2.5×10^6 DO11.10 transgenic T cells intravenously. Spleen dendritic cells cultured for 4 h in the presence of PA alone (200 ng ml⁻¹), PA (200 ng ml⁻¹) plus LF (10 µg ml⁻¹), or medium were washed and pulsed with 10 µg ml⁻¹ OVA(323–339) peptide for 1 h, washed, and then injected into the footpads of reconstituted BALB/c mice. Four days later, immune responses in the draining lymph nodes were analysed by flow cytometric analyses of CD4⁺ KJ126⁺ cells. To determine *in vitro* proliferation, single cell suspensions of draining lymph node cells were cultured with various concentrations of OVA(323–339) peptide, in triplicate in 96-well round-bottomed plates (Costar) in 200 µl RPMI complete medium supplemented with 5% FBS. Proliferative responses were assessed by [³H]thymidine labelling after 72 h of culture in a humidified atmosphere of 5% CO₂ in air. Cytokines in culture supernatants were measured by ELISA kits from Pharmingen.

Received 16 March; accepted 9 June 2003; doi:10.1038/nature01794.

- Mourez, M. *et al.* 2001: a year of major advances in anthrax toxin research. *Trends Microbiol.* **10**, 287–293 (2002).
- Dixon, T. C., Meselson, M., Guillemin, J. & Hanna, P. C. Anthrax. *N. Engl. J. Med.* **341**, 815–826 (1999).
- Lane, H. C., Montagne, J. L. & Fauci, A. Bioterrorism: a clear and present danger. *Nature Med.* **7**, 1271–1273 (2001).
- Mock, M. & Fouet, A. Anthrax. *Annu. Rev. Microbiol.* **55**, 647–671 (2001).
- Little, S. F. & Ivins, B. E. Molecular pathogenesis of *Bacillus anthracis* infection. *Microbes Infect.* **1**, 131–139 (1999).
- Hanna, P. C. & Ireland, J. A. Understanding *Bacillus anthracis* pathogenesis. *Trends Microbiol.* **7**, 180–182 (1999).
- Smith, H. Discovery of the anthrax toxin: the beginning of *in vivo* studies on pathogenic bacteria. *Trends Microbiol.* **8**, 199–200 (2000).
- Brossier, F. & Mock, M. Toxins of *Bacillus anthracis*. *Toxicon* **39**, 1747–1755 (2001).
- Bradley, K. A., Mogridge, J., Mourez, M., Collier, R. J. & Young, J. A. Identification of the cellular receptor for anthrax toxin. *Nature* **414**, 225–229 (2001).
- Duesbery, N. S. *et al.* Proteolytic inactivation of MAP-kinase-kinase by anthrax lethal factor. *Science* **280**, 734–737 (1998).
- Vitale, G., Bernardi, L., Napolitani, G., Mock, M. & Montecucco, C. Susceptibility of mitogen-activated protein kinase kinase family members to proteolysis by anthrax lethal factor. *Biochem. J.* **352**, 739–745 (2000).
- Pellizzari, R., Guidi-Rontani, C., Vitale, G., Mock, M. & Montecucco, C. Anthrax lethal factor cleaves MKK3 in macrophages and inhibits the LPS/IFN γ -induced release of NO and TNF α . *FEBS Lett.* **462**, 199–204 (1999).
- Park, J. M., Greden, F. R., Li, Z. W. & Karin, M. Macrophage apoptosis by anthrax lethal factor through p38 MAP kinase inhibition. *Science* **297**, 2048–2051 (2002).
- Dong, C., Davis, R. J. & Flavell, R. A. MAP Kinases in the immune response. *Annu. Rev. Immunol.* **20**, 55–72 (2002).
- Hanna, P. C., Acosta, D. & Collier, R. J. On the role of macrophages in anthrax. *Proc. Natl Acad. Sci. USA* **90**, 10198–10201 (1993).
- Banchereau, J. & Steinman, R. M. Dendritic cells. *Nature* **392**, 245–252 (1998).
- Pulendran, B., Palucka, K. & Banchereau, J. Sensing pathogens and tuning immune responses. *Science* **293**, 253–256 (2001).
- Liu, Y.-J. Dendritic cell subsets and lineages, and their functions in innate and adaptive immunity. *Cell* **106**, 259–262 (2001).
- Shortman, K. & Liu, Y.-J. Mouse and human dendritic cell subtypes. *Nature Rev. Immunol.* **2**, 151–161 (2002).
- Jernigan, D. B. *et al.* Investigation of bioterrorism-related anthrax, United States, 2001: epidemiologic findings. *Emerg. Infect. Dis.* **8**, 1019–1028 (2002).
- Pulendran, B. *et al.* Developmental pathways of dendritic cells *in vivo*: distinct function, phenotype, and localization of dendritic cell subsets in FLT3 ligand-treated mice. *J. Immunol.* **159**, 2222–2231 (1997).
- Sailaja, G. *et al.* Long-term maintenance of gp120-specific immune responses by genetic vaccination with the HIV-1 envelope genes linked to the gene encoding Flt-3 ligand. *J. Immunol.* **170**, 2496–2507 (2003).
- Liu, L. *et al.* Visualization and quantification of T cell-mediated cytotoxicity using cell-permeable fluorogenic caspase substrates. *Nature Med.* **8**, 185–189 (2002).
- Kearney, E. R., Pape, K. A., Loh, D. Y. & Jenkins, M. K. Visualization of peptide-specific T cell immunity and peripheral tolerance induction *in vivo*. *Immunity* **1**, 327–339 (1994).
- Haskins, K. *et al.* The major histocompatibility complex-restricted antigen receptor on T cells. I. Isolation with a monoclonal antibody. *J. Exp. Med.* **157**, 1149–1169 (1983).
- Ingulli, E., Mondino, A., Khoruts, A. & Jenkins, M. K. *In vivo* detection of dendritic cell antigen presentation to CD4(+) T cells. *J. Exp. Med.* **185**, 2133–2144 (1997).
- Yi, A. K. *et al.* Role of mitogen-activated protein kinases in CpG DNA-mediated IL-10 and IL-12 production: central role of extracellular signal-regulated kinase in the negative feedback loop of the CpG DNA-mediated Th1 response. *J. Immunol.* **168**, 4711–4720 (2002).
- Yi, A. K. *et al.* Lipopolysaccharide and CpG DNA synergize for tumor necrosis factor- α production through activation of NF- κ B. *Int. Immunol.* **13**, 1391–1404 (2002).
- Pannifer, A. D. *et al.* Crystal structure of the anthrax lethal factor. *Nature* **414**, 229–233 (2001).
- Petosa, C. *et al.* Crystal structure of the anthrax protective antigen. *Nature* **385**, 833–838 (1997).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank R. Ahmed and S. Mahanty for critical review of this manuscript. This work was supported by grants from the NIH and CDC.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.P. (bpulend@rmy.emory.edu).

Two-piconewton slip bond between fibronectin and the cytoskeleton depends on talin

Guoying Jiang*, Grégory Giannone*, David R. Critchley†, Emiko Fukumoto‡ & Michael P. Sheetz*

* Department of Biological Sciences, Columbia University, 1212 Amsterdam Avenue, New York, New York 11027, USA

† Department of Biochemistry, University of Leicester, University Road, Leicester LE1 7RH, UK

‡ National Institutes of Dental and Craniofacial Research, National Institute of Health, Bethesda, MD 20892, USA

Mechanical forces on matrix–integrin–cytoskeleton linkages are crucial for cell viability, morphology and organ function¹. The production of force depends on the molecular connections for extracellular–matrix–integrin complexes to the cytoskeleton^{2,3}. The minimal matrix complex causing integrin–cytoskeleton connections is a trimer of fibronectin’s integrin-binding domain FNIII7–10 (ref. 4). Here we report a specific, molecular slip bond that was broken repeatedly by a force of 2 pN at the cellular loading rate of 60 nm s⁻¹; this occurred with single trimer beads but not with monomer. Talin1, which binds to both integrins and actin filaments *in vitro*, is required for the 2-pN slip bond and rapid cytoskeleton binding. Further, inhibition of fibronectin binding to $\alpha_5\beta_3$ and deletion of β_3 markedly decreases the 2-pN force peak. We suggest that talin1 initially forms a molecular slip bond between closely packed fibronectin–integrin complexes and the actin cytoskeleton, which can apply a low level of force to fibronectin until many bonds form or a signal is received to activate a force response.

New connections between extracellular matrix and integrin are formed and used as sites for the application of migration force at the leading edge of migratory cells⁵. The integrin-containing focal contacts behave as individual mechanosensors that exhibit directional assembly in response to local force^{6,7}. However, the molecular nature of connections at early times is not clear, particularly between stationary matrices and moving actin filaments. We observed that the soluble trimeric, but not the monomeric, fibronectin domain FNIII7–10 bound specifically to the leading edge in newly spread cells (Supplementary Fig. S1) before being localized to actin fibre bundles⁴. To analyse the binding of single fibronectin trimers, we coated small (0.64-µm) silica beads with fibronectin trimer diluted with BSA, which gave 30–40% bead binding to cells after 3 s of cell contact in the laser tweezers (BSA-coated beads had 19% binding). Assuming a Poisson distribution of trimers on the beads, there was a high probability (more than 80%) that trimer-bound beads were bound by single FNIII7–10 trimer. When FNIII7–10 monomer beads are bound at the 30–40% level, beads are bound by two or more connections between fibronectin monomer and integrin because single bonds are rapidly reversible (D. Choquet, D. P. Felsenfeld and M.P.S., unpublished observations). Thus, conditions were chosen to have similar numbers

of fibronectin–integrin bonds for the monomer and single trimer beads.

The laser-tweezers microscope enabled us to characterize the force-dependent behaviour of these minimal molecular complexes. When a bead coated with a low density of fibronectin trimer bound to a motile lamellipodium, the bead moved rearward at about 60 nm s^{-1} , which exerted force on the linkage from the bead to the cytoskeleton. Because the initial connections to fibronectin beads were rather weak⁸, we expected that the connections would break at relatively small forces. Beads continued to move out of the trap until the breaking force was reached, and then abruptly returned to the centre of the laser trap as a result of the breakage of the cytoskeleton linkage. This kind of ‘breaking event’ (Fig. 1a) typically occurred in 3 s and within 500 nm of the trap centre (the limit of the linear force–distance relationship of the laser tweezers).

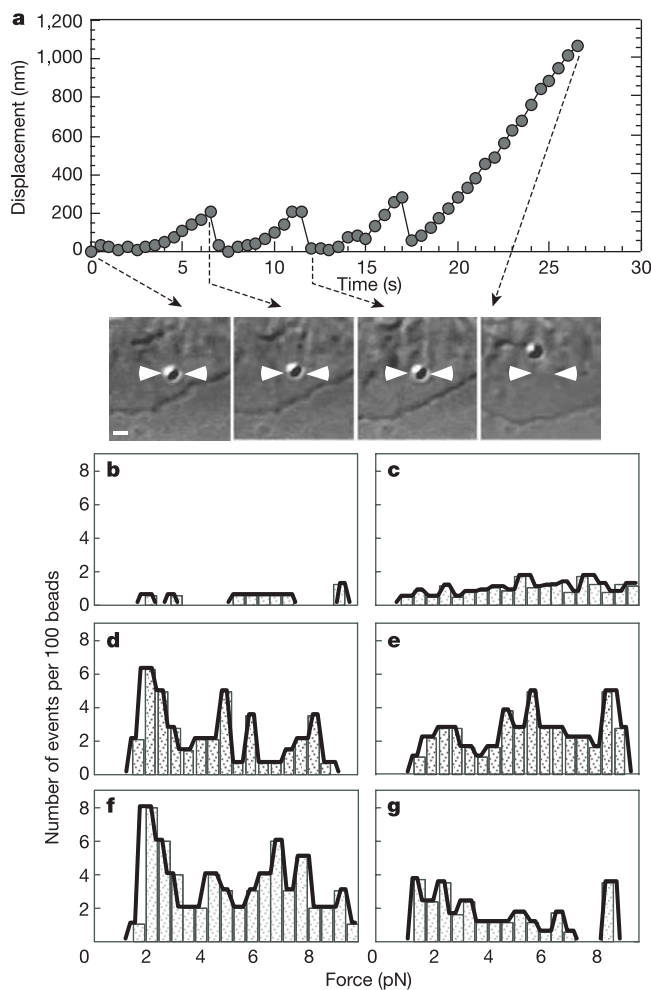


Figure 1 A discrete fibronectin trimer– $\alpha_v\beta_3$ integrin linkage was broken by 2 pN. **a**, Plot of the displacement of a cell-bound fibronectin-trimer-coated bead from the trap centre against time (upper); sequential DIC images of the bead, showing bead movement and the breaking events (lower). The arrows in the lower panels indicate the position of the laser trap. Scale bar, 500 nm. Movies of the ‘breaking events’ can be viewed in Supplementary Information. In RPTP $\alpha^{+/+}$ cells, beads coated with BSA (**b**) (14 breaking events, 153 beads), fibronectin monomer (**c**) (161 breaking events, 740 beads), fibronectin trimer (**d**) (128 breaking events, 315 beads), fibronectin trimer in the presence of Gpen (**e**) (101 breaking events, 402 beads), fibronectin-trimer-coated beads in β_3 wild-type cells (**f**) (99 breaking events, 153 beads) and fibronectin-trimer-coated beads in β_3 knockout cells (**g**) (34 breaking events, 122 beads). Lines enveloping the distributions are the smoothing plots generated by S-PLUS. The stiffness of the laser trap was $40 \text{ pN } \mu\text{m}^{-1}$.

The maximum linkage strength for each ‘breaking event’ was measured from the maximum displacement of the bead from the trap centre when the bond to the cytoskeleton was broken. In histogram plots of the breaking forces, we found that the weakest linkage to the cytoskeleton was commonly broken by 2 pN force with FNIII7-10 trimer-coated beads ($P = 0.85$) (Fig. 1d), whereas beads coated with FNIII7-10 monomer showed no discrete peak ($P = 0.44$) (Fig. 1c). Non-specific binding of BSA contributed little background noise (Fig. 1b). As indicated in Table 1, on average for 100 beads tested, only 19 BSA beads bound to the cell surface and, of these, about 7 (39%) produced 10 breaking events. However, of 100 fibronectin trimer-coated beads, 35 bound to cells and 13 (37%) produced a total of 40 breaking events. For fibronectin monomer-coated beads that bound similarly (39%), there was a slightly higher breaking percentage (41%), but they produced fewer breaking events (20) per 100 beads. Trimer-coated beads usually ($\sim 73\%$) produced more than one breaking event (Table 1), whereas monomer-coated beads primarily produced one ($\sim 63\%$). We also found that the attachment to the cytoskeleton and rearward movement was significantly more rapid with trimer beads ($1.9 \pm 0.3 \text{ s}$ for 100 nm displacement) than monomer-coated beads ($3.4 \pm 0.3 \text{ s}$) or BSA-coated beads ($3.6 \pm 0.4 \text{ s}$).

Fibronectin can bind to both $\alpha_v\beta_3$ and $\alpha_5\beta_1$ integrins on the cell surface⁹. To address which linkage was broken by the 2-pN force, we added the cyclic peptide GPenGRGDSPCA (GPen, where Pen stands for penicillamine), a selective, competitive inhibitor of the vitronectin receptor ($\alpha_v\beta_3$)¹⁰. As shown in Fig. 1e, the 2-pN force peak was markedly decreased ($P = 0.59$) when bead binding to $\alpha_v\beta_3$ was inhibited. Similar results were observed in β_3 knockout cells, in which the 2-pN peak was largely eliminated (Fig. 1g) compared with the control cells (Fig. 1f). Taken together, the above data indicate

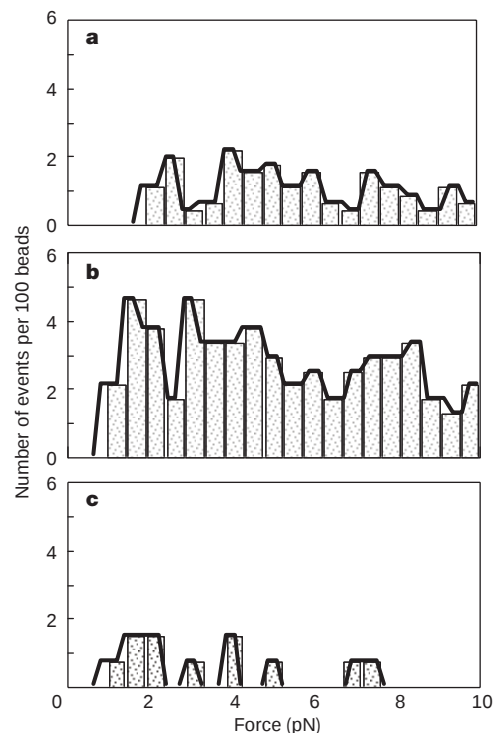


Figure 2 The 2-pN force peak requires talin1. **a**, Talin1 $^{-/-}$ cells (140 breaking events, 456 beads); **b**, talin1 $^{-/-}$ wild-type cells (142 breaking events, 238 beads); **c**, talin1 $^{-/-}$ ABS cells (11 breaking events, 135 beads). Lines enveloping the distributions are the smoothing plots generated by S-PLUS. The stiffness of the laser trap was $40 \text{ pN } \mu\text{m}^{-1}$.

Table 1 Comparison of breaking events between BSA, fibronectin trimer and fibronectin monomer–integrin linkage

Bead coating	Binding (%)	Breaking (%)	No. of breaking events per 100 beads	No. of breaking events per bead
BSA	19 ± 2	39 ± 6	10 ± 4	0.8 ± 0.3
Fibronectin monomer	39 ± 5	41 ± 5	20 ± 5	1.3 ± 0.2
Fibronectin trimer	35 ± 3	37 ± 1	40 ± 5	3.1 ± 0.6

Results shown are means ± s.e.m. for three independent experiments.

that the 2-pN force peak was contributed mostly by the dynamic formation of connections between fibronectin and $\alpha_v\beta_3$ integrin.

There are several reasons to believe that the break occurs between integrins and the cytoskeleton instead of between fibronectin and integrins. The force required to break a single fibronectin–integrin bond is about 20 pN (ref. 11) or about 100 pN to break a single fibrinogen–integrin linkage¹². We have multiple bonds linking each bead to the cell and we break the bead–cytoskeleton linkage with 2 pN. This is even lower than the force generated by a single myosin molecule (4–5 pN)¹³ and the force needed to break a rigor myosin–actin bond (7 pN)¹⁴. This small breaking force is comparable to the breaking force for beads bound to glycosylphosphatidylinositol-linked major histocompatibility complex molecular aggregates that were pulled laterally through the membrane¹⁵. Common to both complexes is the binding to actin filaments.

The actin-binding protein, talin1, binds to the cytoplasmic tails of integrins and is thought to connect integrins with the cytoskeleton². When breaking events were measured in talin1^{-/-} cells, no 2-pN discrete peak was observed ($P = 0.28$) (Fig. 2a). However, a 2-pN force peak was restored ($P = 0.91$) in cells transiently expressing the full-length talin1 complementary DNA (talin1^{-/-} wild-type cells) (Fig. 2b). Expression of a talin1 cDNA lacking the

coding sequence for the carboxy-terminal actin-binding site (talin1^{-/-} ABS cells) markedly decreased the number of breaking events per 100 beads (Fig. 2c; Supplementary Table 1), which implied that the formation of the 2-pN complex was compromised when the linkage between talin and the actin cytoskeleton was disrupted. Both full-length talin1 and talin1-ABS were expressed at the same level, as detected by western blotting (data not shown). Talin2, another talin immunoreactive protein, was also detected in talin1^{-/-} cells (D. H. Sutton, S. W. Craig and D.R.C., unpublished observations). However, it could not compensate for the loss of talin1 in adhesion site formation and stabilization¹⁶, nor does talin2 in these cells replace talin1 in the formation of the 2-pN linkage.

Talin1 markedly increased the rate of bead attachment to the cytoskeleton. Typically, it took 10–13 s for a bound bead to move 50 nm from the trap centre on talin1^{-/-} cells compared with 2–5 s for the same bead on talin1^{-/-} wild-type cells¹⁶. In talin1^{-/-} ABS cells, most (68%) of the bound beads could not move out of the trap. Bead binding was markedly reduced on the talin1^{-/-} and talin1^{-/-} ABS cells compared with talin1^{-/-} wild-type transfectants (27% and 32% versus 47%), respectively (Supplementary Table 1). The fact that the FNIII7-10 trimer beads behaved like control BSA beads on talin1^{-/-} and talin1^{-/-} ABS cells might reflect the fact that the specific binding was a small fraction of the bound beads (sample minus BSA control; 27% – 19% = 8%). The decreased binding was not due to a decrease in integrins on the surface (D. H. Sutton, A. J. Woods, J. C. Norman and D.R.C., unpublished observations); however, the loss of talin might lead to the inactivation of surface integrins, because the talin head, which contains the integrin-binding site, was recently shown to activate $\alpha_{IIb}\beta_3$ integrin¹⁷. In a possibly related observation, talin was shown to act by stably linking multiple extracellular-matrix-bound integrins to the cytoskeleton¹⁸, which could change the avidity of binding¹⁹.

When cells generate force on matrix ligands, they respond to that force by strengthening the connections between the liganded integrins and the force-generating cytoskeleton, in a process defined as ‘reinforcement’⁸ that involves the recruitment of vinculin to the bead⁷. If reinforcement or the proteins involved in reinforcement, such as vinculin or paxillin, were involved in creating the 2-pN bond with single trimer beads, cells defective in reinforcement and paxillin recruitment would show alterations in the 2-pN bond. Because the transmembrane receptor-like protein-tyrosine phosphatase- α (RPTP α) has been shown to be essential for force-dependent reinforcement correlated with paxillin recruitment²⁰ and normal cell spreading on fibronectin²¹, RPTP α ^{-/-} and RPTP α ^{+/+} control cells were tested with fibronectin trimer beads. The same 2-pN peak was always observed (Fig. 3b, c) and talin1 was recruited normally (data not shown). We therefore suggest that the 2-pN bond represents the strength of the initial connection formed between fibronectin–integrin–talin complexes and the cytoskeleton before the reinforcement or recruitment of vinculin or paxillin.

The binding of fibronectin trimer, not the monomer, causes the attachment of integrins to the cytoskeleton at the leading edge of a spreading cell (Supplementary Fig. 1). These connections form the elemental complexes, which might be reinforced or remodelled in focal contact formation and cell motility. We speculate that the complex giving the 2-pN breaking events is minimal and specific for

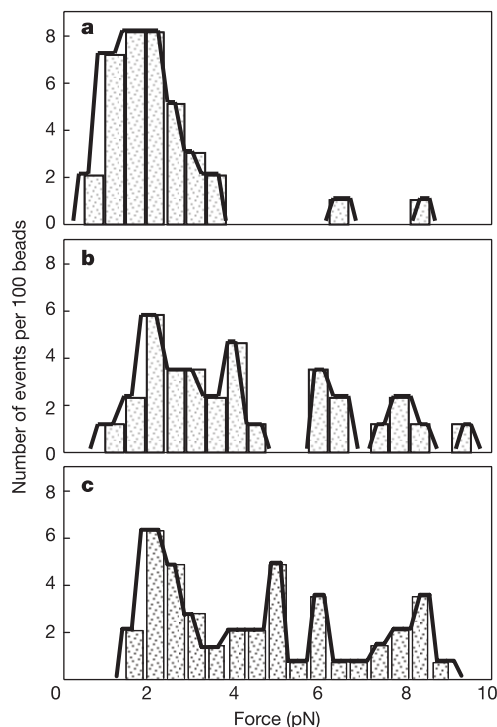


Figure 3 The 2-pN force peak was observed in three different cell lines. **a**, NIH 3T3 fibroblast cells (74 breaking events, 196 beads); **b**, RPTP α ^{-/-} cells (62 breaking events, 172 beads); **c**, RPTP α ^{+/+} cells (128 breaking events, 315 beads). Lines enveloping the distributions are the smoothing plots generated by S-PLUS. The stiffness of the laser trap was 40 pN μm^{-1} .

the following reasons: first, the 2-pN force is the lowest ever observed without the involvement of reinforcement (Fig. 3b); second, different cell lines reliably show the 2-pN peak (Fig. 3a–c); third, the deletion of talin1 specifically removes the 2-pN peak (Fig. 2a), and last, an inhibitor of fibronectin- $\alpha_v\beta_3$ binding or the deletion of β_3 integrin (Fig. 1e, g) markedly decreases the 2-pN peak. We therefore suggest that a minimal complex of three $\alpha_v\beta_3$ integrins, a talin1 dimer and actin filaments connects a single FNIII7-10 trimer to the cytoskeleton. In comparison with other measurements of integrin–cytoskeleton bond strengths²², this 2-pN fibronectin–integrin complex has a strength of about 1 pN per fibronectin–integrin ($\alpha_v\beta_3$)–cytoskeleton linkage, which is similar to the value obtained in ref. 22.

The stronger connections of 4 and 8 pN could be multiples of the basic 2-pN connections; however, they were not observed for some cell lines and there was variability in the pattern. Larger-force peaks existed with all kinds of beads (Fig. 1b–d) and regardless of the presence of β_3 integrin and talin (Figs 1f, g and 2a, b). These connections might therefore involve ‘non-specific’ interactions between the cytoskeleton and the membrane protein aggregates, because the rearward movement of the cytoskeleton past aggregates could cause attachments to form, the breaking of which requires larger forces.

Many observations of single-molecule mechanics indicate that mechanical changes are crucial for protein function^{23–26} and it is expected that mechanical factors can control molecular events in living cells. We suggest that the spacing between liganded integrins is important for recruitment of talin (a single talin molecule could bridge up to four integrins^{2,27}) and that those minimal complexes serve as a slip bond that permits the transfer of a force of about 1 pN per integrin molecule from the actin cytoskeleton to fibronectin. When many minimal complexes are formed in a region, they can serve as a scaffold for the further assembly of focal contacts and for higher-order processes that are crucial to development. □

Methods

Cells, transfection and plasmids

NIH 3T3 mouse fibroblasts, β_3 -null and wild-type cells²⁸, RPTP $\alpha^{+/+}$ cells and RPTP $\alpha^{-/-}$ cells²⁹ were cultured in DMEM medium supplemented with 10% (v/v) fetal bovine serum. The talin $^{-/-}$ ‘fibroblast-like’ cell line²⁹ was maintained in DMEM/F-12 medium supplemented with 15% (v/v) fetal bovine serum (Invitrogen). cDNAs encoding haemagglutinin-tagged full-length talin1 and talin lacking the C-terminal actin-binding sites (talin1 ABS; residues 1–2229) were transiently co-expressed with green fluorescent protein with the use of the Eugene 6 transfection reagent (Roche).

Preparation of biotinylated ovalbumin silica beads

One hundred microlitres of 0.64- μ m silica beads (Bangs Laboratories) were washed three times in 0.2 mM borate buffer pH 8.0 before being suspended in 500 μ l 2 M sodium carbonate buffer (pH 12.0). The suspended beads were incubated at 25 °C with 50 μ l cyanogen bromide (2 mg ml⁻¹ in acetonitrile (Sigma)) for 2 min, and then immediately diluted in 10 ml ice-cold water and centrifuged for 10 min at 2,500g. The beads were further washed twice in 10 ml ice-cold 0.2 mM borate buffer pH 8.0, suspended in 500 μ l 0.2 mM borate buffer pH 8.0 and incubated for 24 h with 500 μ l 4 mg ml⁻¹ ovalbumin (Sigma) (in borate buffer) at 4 °C. The beads were then washed and suspended in 500 μ l PBS buffer pH 7.0, and incubated for 1 h with 15 μ l 30 mg ml⁻¹ N-hydroxysuccinimido-LC-biotin (Pierce) at 25 °C. The biotinylated ovalbumin beads were washed and stored in 0.1% PBS-BSA at 4 °C.

Assay of breaking events

The biotinylated ovalbumin beads were coated with avidin neutralite (150 μ g per 50 μ l of beads) before being coupled overnight with biotinylated fibronectin monomer or trimer in the presence of biotinylated BSA (Sigma) at 4 °C. Beads (1 μ l) were added to each coverslip plated with newly spreading cells.

Cells were plated on acid-washed, silanized coverslips coated with laminin^{8,30}. With an optical-gradient laser trap as described⁸, fibronectin-coated beads were held at the leading edge of the lamellipodia for 3 s, and then released by turning off the laser power. If the bead bound to the cell membrane, the laser was turned on again. The movement of the bead restrained by the laser trap was recorded until the bead was finally pulled out of the trap. The images of the beads were digitized with an O2 SGI (Silicone Graphics Inc.) and tracked with the nanotrack program^{8,30}. Only a displacement smaller than 500 nm (roughly the radius of the trap force field) was counted as a ‘breaking event’.

Statistical analysis

To detect the existence of the 2-pN peak, we applied a normal approximation to a binomial test to all histogram plots, as suggested by the statistics department of Columbia University. The derived *P* value indicates the probability of the 2-pN peak's occurring within the bin range of 0.5–2.5 pN.

Received 11 February; accepted 27 May 2003; doi:10.1038/nature01805.

- Geiger, B. & Bershadsky, A. Assembly and mechanosensory function of focal contacts. *Curr. Opin. Cell Biol.* **13**, 584–592 (2001).
- Critchley, D. R. Focal adhesions—the cytoskeletal connection. *Curr. Opin. Cell Biol.* **12**, 133–139 (2000).
- Sheetz, M. P., Felsenfeld, D. P. & Galbraith, C. G. Cell migration: Regulation of force on extracellular-matrix–integrin complexes. *Trends Cell Biol.* **8**, 51–54 (1998).
- Coussen, F., Choquet, D., Sheetz, M. P. & Erickson, H. P. Trimers of the fibronectin cell adhesion domain localize to actin filament bundles and undergo rearward translocation. *J. Cell Sci.* **115**, 2581–2590 (2002).
- Lauffenburger, D. A. & Horwitz, A. F. Cell migration: A physically integrated molecular process. *Cell* **84**, 359–369 (1996).
- Riveline, D. *et al.* Focal contacts as mechanosensors: Externally applied local mechanical force induces growth of focal contacts by an mDia1-dependent and ROCK- independent mechanism. *J. Cell Biol.* **153**, 1175–1186 (2001).
- Galbraith, C. G., Yamada, K. M. & Sheetz, M. P. The relationship between force and focal complex development. *J. Cell Biol.* **159**, 695–705 (2002).
- Choquet, D., Felsenfeld, D. P. & Sheetz, M. P. Extracellular matrix rigidity causes strengthening of integrin–cytoskeleton linkages. *Cell* **88**, 39–48 (1997).
- Sonnenberg, A. Integrins and their ligands. *Curr. Top. Microbiol. Immunol.* **184**, 7–35 (1993).
- Pierschbacher, M. D. & Ruoslahti, E. Influence of stereochemistry of the sequence Arg-Gly-Asp-Xaa on binding specificity in cell adhesion. *J. Biol. Chem.* **262**, 17294–17298 (1987).
- Thoumine, O., Kocian, P., Kottelat, A. & Meister, J. J. Short-term binding of fibroblasts to fibronectin: Optical tweezers experiments and probabilistic analysis. *Eur. Biophys. J.* **29**, 398–408 (2000).
- Litvinov, R. I., Shuman, H., Bennett, J. S. & Weisel, J. W. Binding strength and activation state of single fibrinogen–integrin pairs on living cells. *Proc. Natl Acad. Sci. USA* **99**, 7426–7431 (2002).
- Finer, J. T., Mehta, A. D. & Spudich, J. A. Characterization of single actin–myosin interactions. *Biophys. J.* **68**, 291S–296S (1995).
- Nishizaka, T., Miyata, H., Yoshikawa, H., Ishiwata, S. & Kinoshita, K. Jr Unbinding force of a single motor molecule of muscle measured using optical tweezers. *Nature* **377**, 251–254 (1995).
- Suzuki, K. & Sheetz, M. P. Binding of cross-linked glycosylphosphatidylinositol-anchored proteins to discrete actin-associated sites and cholesterol-dependent domains. *Biophys. J.* **81**, 2181–2189 (2001).
- Giannone, G., Jiang, G., Sutton, D. H., Critchley, D. R. & Sheetz, M. P. Talin1 is essential for force-dependent reinforcement of integrin–cytoskeleton connections but not for activation of tyrosine kinases. *J. Cell Biol.* (submitted).
- Calderwood, D. A. *et al.* The Talin head domain binds to integrin beta subunit cytoplasmic tails and regulates integrin activation. *J. Biol. Chem.* **274**, 28071–28074 (1999).
- Brown, N. H. *et al.* Talin is essential for integrin function in *Drosophila*. *Dev. Cell* **3**, 569–579 (2002).
- Nishizaka, T., Shi, Q. & Sheetz, M. P. Position-dependent linkages of fibronectin–integrin–cytoskeleton. *Proc. Natl Acad. Sci. USA* **97**, 692–697 (2000).
- von Wichert, G. *et al.* RPTP- α acts as a transducer of mechanical force on $\alpha_v\beta_3$ –integrin–cytoskeleton linkages. *J. Cell Biol.* **161**, 143–153 (2003).
- Su, J., Muranjan, M. & Sap, J. Receptor protein tyrosine phosphatase α activates Src-family kinases and controls integrin-mediated responses in fibroblasts. *Curr. Biol.* **9**, 505–511 (1999).
- Balaban, N. Q. *et al.* Force and focal adhesion assembly: A close relationship studied using elastic micropatterned substrates. *Nature Cell Biol.* **3**, 466–472 (2001).
- Watanabe, K. *et al.* Molecular mechanics of cardiac titin's PEVK and N2B spring elements. *J. Biol. Chem.* **277**, 11549–11558 (2002).
- Oberhauser, A. F., Marszalek, P. E., Erickson, H. P. & Fernandez, J. M. The molecular elasticity of the extracellular matrix protein tenascin. *Nature* **393**, 181–185 (1998).
- Carrión-Vázquez, M. *et al.* Mechanical and chemical unfolding of a single protein: A comparison. *Proc. Natl Acad. Sci. USA* **96**, 3694–3699 (1999).
- Merkel, R., Nassoy, P., Leung, A., Ritchie, K. & Evans, E. Energy landscapes of receptor–ligand bonds explored with dynamic force spectroscopy. *Nature* **397**, 50–53 (1999).
- Xing, B., Jedsadayamata, A. & Lam, S. C. Localization of an integrin binding site to the C terminus of talin. *J. Biol. Chem.* **276**, 44373–44378 (2001).
- Hodivala-Dilke, K. M. *et al.* β_3 -integrin-deficient mice are a model for Glanzmann thrombasthenia showing placental defects and reduced survival. *J. Clin. Invest.* **103**, 229–238 (1999).
- Pridle, H. *et al.* Disruption of the talin gene compromises focal adhesion assembly in undifferentiated but not differentiated embryonic stem cells. *J. Cell Biol.* **142**, 1121–1133 (1998).
- Felsenfeld, D. P., Choquet, D. & Sheetz, M. P. Ligand binding regulates the directed movement of β_1 integrins on fibroblasts. *Nature* **383**, 438–440 (1996).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank H. Erickson for FNIII7-10 monomer and trimer constructs; A. J. Woods for the talin1-ABS construct; R. Hynes for β_3 cell lines; J. Sap for RPTP α cell lines; G. von Wichert and N. Heckenberg for discussions; and all the Sheetz laboratory members for consistent help. This work was supported by a grant from the NIH to M.P.S. Work in D.R.C.'s laboratory was funded by the Wellcome Trust.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.P.S. (ms2001@columbia.edu).

Structural transitions and elasticity from torque measurements on DNA

Zev Bryant*, Michael D. Stone*, Jeff Gore†, Steven B. Smith†‡, Nicholas R. Cozzarelli* & Carlos Bustamante*†§

* Department of Molecular and Cell Biology, † Department of Physics, ‡ Howard Hughes Medical Institute, and § Physical Biosciences Division of Lawrence Berkeley National Laboratory, University of California, Berkeley, California 94720, USA

Knowledge of the elastic properties of DNA is required to understand the structural dynamics of cellular processes such as replication and transcription. Measurements of force and extension on single molecules of DNA^{1–3} have allowed direct determination of the molecule's mechanical properties, provided rigorous tests of theories of polymer elasticity⁴, revealed unforeseen structural transitions induced by mechanical stresses^{3,5–7}, and established an experimental and conceptual framework for mechanical assays of enzymes that act on DNA⁸. However, a complete description of DNA mechanics must also consider the effects of torque, a quantity that has hitherto not been directly measured in micromanipulation experiments. We have measured torque as a function of twist for stretched DNA—torsional strain in over- or underwound molecules was used to power the rotation of submicrometre beads serving as calibrated loads. Here we report tests of the linearity of DNA's twist elasticity, direct measurements of the torsional modulus (finding a value ~40% higher than generally accepted), characterization of torque-induced structural transitions, and the establishment of a framework for future assays of torque and twist generation by DNA-dependent enzymes. We also show that cooperative structural transitions in DNA can be exploited to construct constant-torque wind-up motors and force–torque converters.

Previous investigations of the force–extension behaviour of supercoiled DNA have found that, under low tension, DNA behaves like an isotropic flexible rod^{2,9}: as turns are added to the molecule, its extension remains nearly constant until a critical twist density is

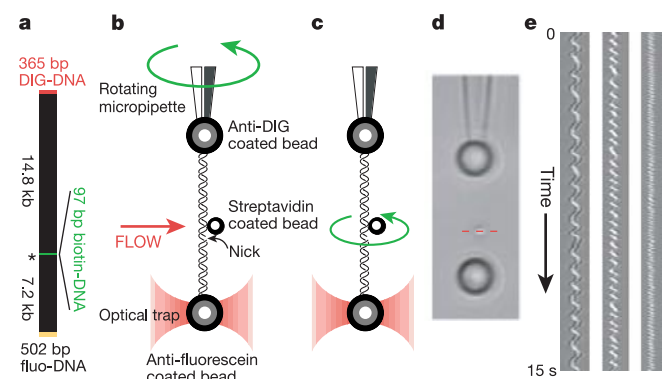


Figure 1 Experimental design. **a**, The molecular construct contains three distinct attachment sites and a site-specific nick (asterisk), which acts as a swivel. **b**, Each molecule was stretched between two antibody-coated beads using a dual-beam optical trap¹¹. A rotor bead was then attached to the central biotinylated patch (see Supplementary Movie 1). The rotor was held fixed by applying a fluid flow, and the micropipette was twisted to build up torsional strain in the upper segment of the molecule. **c**, Upon releasing the flow, the central bead rotated to relieve the torsional strain. **d**, Video of the rotating bead (Supplementary Movies 2 and 3) was analysed to track cumulative changes in angle. Horizontal sections (red dashed line) of successive video frames can be stacked (**e**) to allow visualization of the helical path of the bead in space and time. Left to right, traces of 920 nm, 760 nm and 520 nm rotor beads.

reached and the molecule buckles to form plectonemic (interwound) structures; thereafter, additional turns cause a rapid decrease in extension as twist is traded for writhe. At higher tensions, the behaviour of over- and underwound molecules differ. In each case, DNA undergoes a structural change before the twist density necessary for buckling is reached, and the molecule lengthens (or contracts only gradually) as additional turns are introduced^{6,7}. These results have been interpreted to reflect cooperative torque-induced transitions in DNA structure^{6,7}, and a theoretical force–torque phase diagram has been proposed that explains the major features of force–extension curves for torsionally constrained molecules over a large range of forces and supercoiling densities¹⁰. Our work makes direct tests of both the isotropic rod model and the proposed force–torque phase diagram.

To perform dynamic torque measurements on single DNA molecules, we generated molecular constructs shown diagrammatically in Fig. 1a. The use of three distinct chemical modifications of DNA allows oriented tethering of the ends of the molecule, and attachment of a small bead (the ‘rotor’) to an internal position. A site-specific nick engineered below the rotor attachment point serves as a swivel.

We begin each assay by assembling the DNA molecule and rotor between two antibody-coated beads held in a micropipette and a force-measuring optical trap¹¹ (Fig. 1b and Supplementary Movie 1). Typically, we build up torsional stress in the molecule by holding the rotor bead stationary using fluid flow, and rotating the micropipette by ≥ 300 turns. When the flow is released, torque stored in the upper DNA segment causes the central bead to rotate continuously about its edge (Fig. 1c–e, and Supplementary Movies 2 and 3) until the torsional stress has been relieved, after which the bead rotates slightly back and forth under the influence of thermal fluctuations. Constant tension is maintained using force feedback¹¹, preventing buckling of the molecule² during the experiment, so that the observed dynamics reflect changes in twist and not writhe. At low Reynolds numbers, the magnitude of the torque can be determined as the product of the angular velocity (ω) of the rotor and its rotational drag (γ).

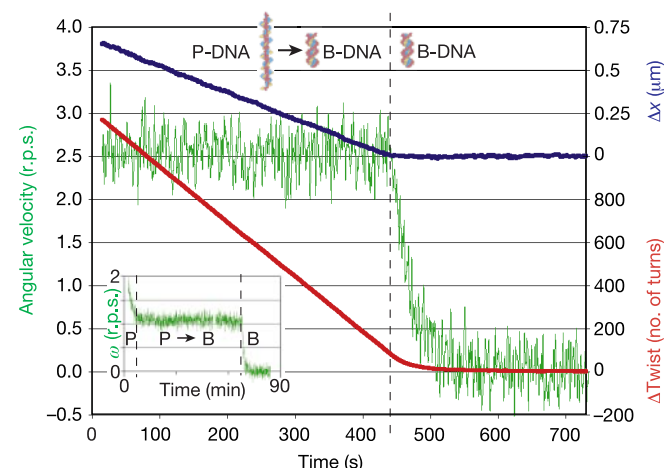


Figure 2 Analysis of bead rotation. A tether was overwound by 1,200 turns before releasing the 760 nm rotor bead. Red, cumulative rotor angle ($\Delta\theta$); green, angular velocity; and blue, molecular extension (Δx) versus time under constant tension (45 pN). The data support a model in which overwinding the molecule converts a fraction of the DNA into P-form^{6,7}. P-DNA converts back into B-form at constant torque, after which the molecule regains its B-form extension and the remaining torque decays to zero. Inset, a molecule was completely converted to P-DNA by introducing 4,800 turns before releasing the 920 nm rotor bead. Torsional relaxation of hyperwound P-DNA preceded the long (~4,000 turns) P–B transition. During the gap at ~50 min, rotation was paused by turning on flow.

In highly overwound DNA held at tensions >7 pN, a portion of the molecule is converted from standard B-DNA into an overextended, high-helicity form called P-DNA^{6,7}. If the B $\rightarrow$ P transition is highly cooperative, it should occur at a constant torque. Confirming this prediction, the angular velocity of the rotor remained constant for much of the torsional relaxation of overwound molecules (Fig. 2). P to B conversion was reflected in a progressive reduction in extension. When all of the molecule had been converted to B-DNA, the extension ceased to change and the angular velocity of the rotor decayed to zero as the remaining torsional strain in the B-DNA was removed. Complete conversion of the 14.8 kilobase (kb) molecule into P-DNA (Fig. 2, inset) required the introduction of $\sim 4.0 \times 10^3$ turns, implying a helical repeat of ~ 2.7 base pairs (bp) per turn for P-DNA, close to previous estimates^{6,7} of ~ 2.6 bp per turn. From the changes in extension at 45 pN, we conclude that P-DNA is 50% longer than B-DNA under our conditions (slightly shorter than previous estimates^{6,7} of 60–75% longer than B-form).

To quantify the torques (τ) in our experiments, the angular velocity of the rotor during the P $\rightarrow$ B transition ($\omega_{P \rightarrow B}$) was measured using rotor beads of several different radii (r) (Fig. 3a, inset). The intended rotor geometry (a sphere rotating about its edge) was confirmed by video analysis, which showed that the orbital radius was equal to the bead radius (within a 2% error). This geometry predicts $\tau = \gamma\omega = 14\pi r^3\eta\omega$, where η is the viscosity of the solution. The appearance of the expected relationship $\omega \propto r^{-3}$ allowed us to determine the critical torque of the P $\rightarrow$ B transition, $\tau_{crit,+} = 34 \pm 2$ pN nm (at 45 pN tension). This value was thereafter used as a calibration factor: instantaneous angular velocities of individual rotor beads were converted into torques using $\tau = \tau_{crit,+}(\omega/\omega_{P \rightarrow B})$.

To precisely determine the twist elasticity of DNA, numerous molecules were repeatedly over- or underwound and allowed to

relax. Plots of torque as a function of twist show considerable run-to-run variation due to random thermal fluctuations (Fig. 3a). Therefore, torque–twist data from 103 runs were averaged together (Fig. 3b). The torque–twist relationship for DNA shows two constant-torque regions, reflecting structural transitions at $\tau_{crit,+} = +34$ pN nm and $\tau_{crit,-} = -9.6$ pN nm, and a nearly linear region reflecting the twist elasticity of B-DNA. Our direct measurements of $\tau_{crit,+}$ and $\tau_{crit,-}$ are in good agreement with previous estimates from force–extension analysis^{6,7,9,12}. Entry into the 34 pN nm plateau can now be seen to be very sharp, demonstrating the cooperativity of the B–P transition. (Using Ising/Zimm-Bragg theory, we find that a free-energy penalty of at least $4k_B T$ at P–B domain boundaries is required to explain the data.) The transition at -9.6 pN nm is somewhat less cooperative (showing a shallow slope in the initial portion of the plateau), and probably reflects the formation of denatured DNA⁶ and some combination of non-canonical structures (for example, Z-DNA^{7,10}). Because the DNA has a net left-handed twist (~ 13 bp per turn; data not shown) upon completion of this transition^{7,8,10}, we have collectively designated the underwound states ‘L-DNA’.

The twist elasticity of B-DNA is often assumed to be that of an isotropic rod: torque builds up linearly with twist according to $\tau = (C/L)(\theta - \theta_0)$, where L is the rod length, θ is the twisting angle, θ_0 is the equilibrium twist (35° per bp), and the torsional modulus C is constant. However, the asymmetry of twisting a chiral molecule suggests that this approximation may break down. Indeed, a study of fluorescence polarization anisotropy decay (FPA) reported a linear dependence of C on supercoiling density¹³. The anharmonic model proposed from FPA¹³ fits our data better than a simple harmonic model (Fig. 3b). However, the deviation from linearity is small, accounting for a $<10\%$ reduction in torque near entry into the P–B transition.

The torsional modulus C of DNA (near $\theta - \theta_0 = 0$) has been

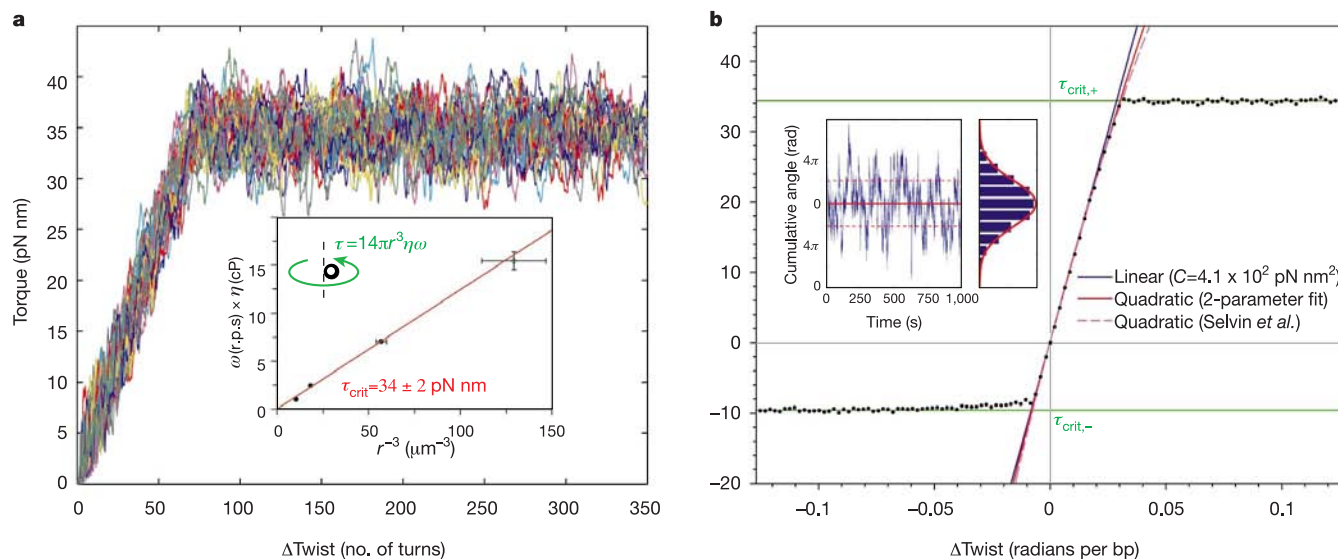


Figure 3 Torque calibration and twist elasticity of DNA. **a**, Torque versus twist for 37 runs from 16 molecules (520 nm rotor beads; 45 pN constant tension). Inset, calibration of $\tau_{crit,+}$. Angular velocities (ω) of beads with mean diameters of 920 nm ($n = 5$), 760 nm ($n = 6$) or 520 nm ($n = 16$) were measured during the P $\rightarrow$ B transition at 45 pN (filled circles, $\pm$ s.d.). Viscosity-corrected ω is proportional to r^{-3} ; the best-fit slope (red) gives $\tau_{crit,+} = 34 \pm 2$ pN nm. Open circle, extrapolated P $\rightarrow$ B angular velocity for 400 nm beads ($n = 3$; see Methods). **b**, Averaged twist elasticity data. Negative torques: average of 39 runs at 15 pN. Positive torques: 37 runs at 45 pN and 27 runs at 15 pN gave very similar traces and were averaged together. Green lines, constant-torque structural transitions. Blue, linear fit to the data points falling within ± 8 pN nm. Anharmonic models

($C(\Delta\theta) = a - b\Delta\theta/N$, where $N = 14,795$ bp) give superior fits to the data over the full range of B-DNA stability. Red, two-parameter anharmonic fit ($b/a = 4.5$); dashed purple, anharmonic fit constraining $b/a = 8.16$ to agree with FPA data¹³. Blue and red fits give $C(0) = 4.1 \times 10^2$ pN nm²; purple fit gives $C(0) = 4.3 \times 10^2$ pN nm². Fits to 45 pN or 15 pN positive torque data alone (not shown) give $C(0) = 4.1 \times 10^2$ or 4.3×10^2 pN nm², respectively. Inset, independent measure of $C(0)$ using the equipartition theorem $\langle \Delta\theta^2 \rangle = k_B TL/C$, where $L = 5.03 \mu m$. Twist fluctuations of five molecules (520 nm rotor beads, 15 pN tension) were tracked at 50 Hz (example shown). Angular variance (red gaussian plot over blue data histogram) gives $C(0) = (4.4 \pm 0.4) \times 10^2$ pN nm².

estimated from FPA^{14,15}, equilibrium topoisomer distribution of ligated circles^{16,17}, and circularization kinetics¹⁸. These methods give a wide range of values for C . FPA typically gives $C \approx 200 \text{ pN nm}^2$, although elevated values have been reported for small circular DNA¹⁵. Topoisomer distribution analysis yields $C = 300\text{--}400 \text{ pN nm}^2$, and best fits as high as 480 pN nm^2 have been reported from circularization kinetics¹⁸. Theoretical treatments of force and extension data from single supercoiled DNA molecules^{9,12,19,20} have yielded $C = 300 \text{ pN nm}^2$ (ref. 19), 350 pN nm^2 (refs 9, 12) and $C = 450 \text{ pN nm}^2$ (ref. 20), depending on the analytical method. For all of the methods above, analytical theories or simulations are needed to estimate the relative contributions of twisting and bending or writhing to the experimental results. Topoisomer distributions of small circles (yielding $C \approx 300 \text{ pN nm}^2$) have generally been considered the most reliable measurements, since the writhing contribution is thought to be small.

Our measurements provide a direct determination of C in the absence of writhing. A simple linear fit to the torque–twist data near $\theta - \theta_0 = 0$ gives $C = 410 \pm 30 \text{ pN nm}^2$ (Fig. 3b). The uncertainty in this value depends heavily on imprecision in the absolute torque calibration from viscous drag (Fig. 2a, inset), since the relative scatter in the averaged data is small. To provide a drag-independent measure of C and check the calibration of our data, we used our experimental system to observe the amplitude of thermal fluctuations in twist (around $\theta - \theta_0 = 0$) and applied the equipartition theorem, as was previously done for single actin filaments²¹. This analysis (Fig. 3b, inset) yields $C = 440 \pm 40 \text{ pN nm}^2$, in good agreement with our dynamic torque measurements.

These direct measurements of the torsional modulus of stretched DNA agree with the largest estimates of C from other methods, and are 40–50% higher than the widely used value of $\sim 300 \text{ pN nm}^2$. The difference may be due to the presence of tension in our experiments. Stretching may induce structural changes that raise the torsional modulus, analogous to the proposed role of bending strain¹⁵, or it may raise the effective C value owing to elastic coupling between

twisting, bending and stretching (A. Matsumoto and W. Olson, personal communication). No increase in rigidity was observed between 15 pN and 45 pN, so any such effect must be saturated at these forces.

It is also possible that previous studies have not sufficiently accounted for writhing, and consequently underestimated the torsional rigidity. In this case, the only important role of tension is suppression of writhing fluctuations, which would otherwise lead to torsional softening²⁰. Our measurement should then accurately reflect the local torsional modulus of the molecule.

We extended our analysis of torque and twist to investigate the overstretching transition of DNA (Fig. 4a). When DNA is pulled to high forces, it undergoes a cooperative transition at $\sim 65 \text{ pN}$ to a form ('S-DNA') that is 70% longer than B-form^{3,5}. The structure of S-DNA remains the subject of debate²², but it is known to be substantially underwound relative to B-DNA^{7,10}. Therefore, untwisting must accompany overstretching. When our molecule-rotor assemblies are held at tensions above 65 pN, torque is generated that spins the rotor as the B $\rightarrow$ S transition progresses (Fig. 4a and Supplementary Movie 4). The ratio of changes in extension and twist ($\Delta x/\Delta\theta = 3.7 \text{ nm per turn}$) corresponds to a helical repeat of $\sim 33 \text{ bp per turn}$ for S-DNA, slightly smaller than the previously reported value of 37.5 bp per turn ^{7,10}. Relaxation of the molecule below 65 pN allows complete rewinding to B-form twist. Higher tensions produce higher unwinding torques, following a Clausius–Clapeyron-like force–torque coexistence line for the B–S transition. Together with tension-dependent critical torque data for the P–B and L–B transitions, these measurements validate the proposed force–torque phase diagram for stretched and twisted DNA¹⁰ (Fig. 4b).

We have shown that an over- or underwound DNA molecule behaves as a constant-torque wind-up motor capable of repeatedly producing thousands of rotations, and that an overstretched molecule acts as a force–torque converter. (During overstretching at forces above $F_{\text{crit}} \approx 65 \text{ pN}$, the efficiency, ϵ , of converting force–extension work into rotational work is given by the Carnot-like expression $\epsilon = 1 - F_{\text{crit}}/F$.) The production of continuous directed rotation by molecular devices has potential applications in the construction of nanomechanical systems. Previous approaches to this problem have included co-opting the biological motor F1 ATP synthase²³, which produces torques comparable to the P–B transition^{23,24}. DNA-based devices that produce continuous rotation, according to the principles demonstrated here, might be integrated with existing DNA nanotechnology²⁵ using the intrinsic self-assembly properties of DNA.

An important future application of the methods developed here is the measurement of torque generation and changes in twist induced by DNA-associated enzymes. A previous measurement of torque generation by RNA polymerase²⁶ was complicated by buckling of the DNA to form plectonemes. Here, we have established a method for torque measurement that decouples twisting from writhing, with the potential for broad application to biological motors. □

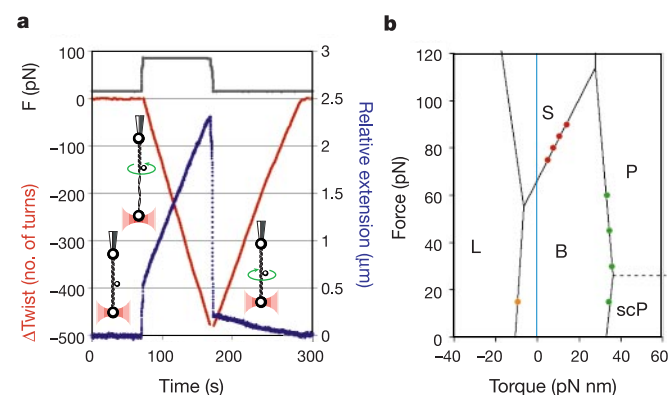


Figure 4 Unwinding during overstretching, and the global force–torque phase diagram of DNA. **a**, Unwinding during overstretching. Constructs were pulled to forces (black trace) above 65 pN. Cumulative rotations (red) of a 400 nm rotor bead showed continuous constant-torque untwisting while tension was maintained at 85 pN (Supplementary Movie 4). Concurrent decrease in twist (red) and increase in extension (blue) reflect S-DNA formation. When the force was relaxed to 15 pN, the molecule rewound, returning to B-form twist and extension. **b**, The global force–torque phase diagram of DNA. A single molecule was unwound by pulling to successively higher forces (relaxing at 15 pN between each run) to measure the torque of the B–S transition (judged from $\omega_B \rightarrow S$) as a function of force (red). The critical torque of the P–B transition was also measured as a function of force by relaxation of overwound tethers under varying tensions (green), and all critical torque data (orange: $\tau_{\text{crit}} = 9.6 \text{ pN nm}$) were found to agree with a simple model¹⁰ in which DNA can access five distinct structural states differing only in extension, twist, and free energy. scP, P-DNA that has been shortened by supercoiling^{6,7}.

Methods

Beads and molecular constructs

Beads (2.8 μm) coated with Protein G (SpheroTech) were crosslinked to rabbit anti-fluorescein (Molecular Probes) or sheep anti-digoxigenin (DIG) (Roche). Streptavidin-coated rotor beads were purchased from SpheroTech (760 nm) or Bangs (all others), and their diameters were measured from transmission electron micrographs using diffraction grating replicas as standards (400 nm, $n = 55$; 520 nm, $n = 81$; 760 nm, $n = 48$; 920 nm, $n = 66$); all beads appeared round.

Molecular constructs were generated by serial ligation of purified restriction fragments of chemically modified and unmodified DNA. Hapten-modified polymerase chain reaction (PCR) fragments (966 bp) were generated by amplification of a multiple cloning site using 0.2 mM dATP, dCTP and dGTP, 0.13 mM dTTP, and 0.07 mM biotin-dUTP, DIG-dUTP or fluorescein-dUTP (Roche). PCR fragments were digested and gel-purified to recover a 374 bp biotin-modified *KpnI* fragment, a 365 bp DIG-modified *BamHI* fragment, and a 502 bp fluorescein-modified *Clal* fragment. pSV8 (a plasmid generated by

insertion of the 5.5 kb *Bam*HI fragment of λ phage into pSV01²⁷) was linearized with *Kpn*I, treated with alkaline phosphatase (AP), and ligated to an excess of the biotinylated *Kpn*I fragment. The product (containing nicked ligation junctions owing to the absence of plasmid-contributed phosphates) was digested with *Cl*al and *Xho*I. The resulting purified 7.3 kb fragment, designated protoSM, has the structure *Cl*al:7.2 kb:nick:97 bp(biotin-modified):*Xho*I. The torque-bearing DNA segment is the 14.8 kb *Bam*HI:*Sal*I fragment of pPIA6²⁸. The final construct (SM1) was generated in a four-way ligation of the fluorescein-modified fragment, protoSM, torque-bearing segment, and DIG-modified fragment. Full-length products were selected by sequential binding to anti-fluorescein- and anti-DIG-coated beads in the optical tweezers. SM1 (Fig. 1a) was used in all reported experiments, with the following exceptions: SM2 (identical except DIG-DNA section extended to 601 bp for added stability) was used for the run shown in Fig. 2. SM3 (torque-bearing segment replaced with the 8.4 kb *Bgl*II-*Sal*I fragment of pSV8; fluorescein-DNA section extended to 4 kb) was used for the force-torque analysis of the B-S transition (red data points in Fig. 4b).

Experimental assembly and data collection

Anti-fluorescein beads were incubated with DNA and introduced into the flow chamber. Anti-DIG beads were introduced via a separate channel, and a molecular tether was assembled by keeping an anti-DIG bead on the micropipette by suction, and 'fishing' near a DNA:anti-fluorescein bead held in the laser trap. The trapped bead was then released into flow, and a streptavidin-coated 'rotor' bead was trapped and brought to the vicinity of the biotinylated portion of the molecule, where it became attached laterally to the DNA (Supplementary Movie 1). The micropipette was rotated using a computer-controlled electric motor (LEGO Mindstorms) while the rotor bead was held fixed by flowing buffer at $\sim 0.5 \text{ mm s}^{-1}$.

All experiments were performed in 100 mM NaCl and 40 mM Tris-HCl (pH 8.2). EDTA was typically present at 1 mM; omission caused no perceptible changes. Ambient temperature ($23 \pm 1^\circ\text{C}$) was recorded prior to each experiment for use in viscosity calculations. Drag was also corrected for hydrodynamic coupling with the outer beads²⁹; correction factors for the different rotor diameters were 1.005 (400 nm), 1.01 (520 nm), 1.02 (760 nm) and 1.03 (920 nm). Video was digitized at 30 Hz unless otherwise indicated, and the instantaneous angle of the rotor was extracted from the *x*-position and brightness (indicative of focal depth) of the bead. Angular velocities were obtained by numerical differentiation of the cumulative bead angle over a 1 s (Fig. 3) or 2 s (Fig. 2) window. The extrapolated $P \rightarrow B$ velocity of 400 nm beads (open circle in Fig. 3a inset) was obtained by measuring the velocity at large negative twists and scaling by $\tau_{\text{crit},+}/\tau_{\text{crit},-}$, since $P \rightarrow B$ rotation was too fast to track. During data collection, constant tension was maintained using stage-based force feedback¹¹. During the exceptionally long run shown in Fig. 2 inset, force feedback (45 pN) was inoperative (out of actuator range) prior to $t = 23$ min, but $F > 30$ pN throughout.

Phase diagram

In the 'zero-temperature' approximation, the five-state structural model¹⁰ leads to force-torque coexistence lines with constant slopes $\delta F/\delta \tau = -\Delta\theta/\Delta x$, where $\Delta\theta$ and Δx are the changes in twist and extension, respectively, for a particular structural transition. The slopes of the boundaries shown (Fig. 4b) were taken from experimental measurements of $\Delta\theta/\Delta x$, and predict the trends of the force-torque measurements. The intercepts of the boundaries were varied to fit the data. No torque measurements were made at the S-L or S-P boundaries, so these predicted slopes remain to be confirmed.

Received 6 February; accepted 28 May 2003; doi:10.1038/nature01810.

- Smith, S. B., Finzi, L. & Bustamante, C. Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. *Science* **258**, 1122–1126 (1992).
- Strick, T. R., Allemand, J. F., Bensimon, D., Bensimon, A. & Croquette, V. The elasticity of a single supercoiled DNA molecule. *Science* **271**, 1835–1837 (1996).
- Smith, S. B., Cui, Y. & Bustamante, C. Overstretching B-DNA: The elastic response of individual double-stranded and single-stranded DNA molecules. *Science* **271**, 795–799 (1996).
- Bustamante, C., Marko, J. F., Siggia, E. D. & Smith, S. Entropic elasticity of lambda-phage DNA. *Science* **265**, 1599–1600 (1994).
- Cluzel, P. *et al.* DNA: An extensible molecule. *Science* **271**, 792–794 (1996).
- Allemand, J. F., Bensimon, D., Lavery, R. & Croquette, V. Stretched and overwound DNA forms a Pauling-like structure with exposed bases. *Proc. Natl Acad. Sci. USA* **95**, 14152–14157 (1998).
- Leger, J. F. *et al.* Structural transitions of a twisted and stretched DNA molecule. *Phys. Rev. Lett.* **83**, 1066–1069 (1999).
- Bustamante, C., Bryant, Z. & Smith, S. B. Ten years of tension: Single-molecule DNA mechanics. *Nature* **421**, 423–427 (2003).
- Bouchiat, C. & Mezdard, M. Elasticity model of a supercoiled DNA molecule. *Phys. Rev. Lett.* **80**, 1556–1559 (1998).
- Sarkar, A., Leger, J. F., Chatenay, D. & Marko, J. F. Structural transitions in DNA driven by external force and torque. *Phys. Rev. E* **63**, 051903 (2001).
- Smith, S. B., Cui, Y. & Bustamante, C. Optical-trap force transducer that operates by direct measurement of light momentum. *Methods Enzymol.* **361**, 134–162 (2003).
- Strick, T. R., Bensimon, D. & Croquette, V. Micro-mechanical measurement of the torsional modulus of DNA. *Genetica* **106**, 57–62 (1999).
- Selvin, P. R. *et al.* Torsional rigidity of positively and negatively supercoiled DNA. *Science* **255**, 82–85 (1992).
- Millar, D. P., Robbins, R. J. & Zewail, A. H. Direct observation of the torsional dynamics of DNA and RNA by picosecond spectroscopy. *Proc. Natl Acad. Sci. USA* **77**, 5593–5597 (1980).
- Heath, P. J., Clendenning, J. B., Fujimoto, B. S. & Schurr, J. M. Effect of bending strain on the torsion elastic constant of DNA. *J. Mol. Biol.* **260**, 718–730 (1996).
- Horowitz, D. S. & Wang, J. C. Torsional rigidity of DNA and length dependence of the free energy of DNA supercoiling. *J. Mol. Biol.* **173**, 75–91 (1984).

- Shore, D. & Baldwin, R. L. Energetics of DNA twisting. II. Topoisomer analysis. *J. Mol. Biol.* **170**, 983–1007 (1983).
- Crothers, D. M., Drak, J., Kahn, J. D. & Levene, S. D. DNA bending, flexibility, and helical repeat by cyclization kinetics. *Methods Enzymol.* **212**, 3–29 (1992).
- Vologodskii, A. V. & Marko, J. F. Extension of torsionally stressed DNA by external force. *Biophys. J.* **73**, 123–132 (1997).
- Moroz, J. D. & Nelson, P. Entropic elasticity of twist-storing polymers. *Macromolecules* **31**, 6333–6347 (1998).
- Yasuda, R., Miyata, H. & Kinoshita, K. Jr Direct measurement of the torsional rigidity of single actin filaments. *J. Mol. Biol.* **263**, 227–236 (1996).
- Williams, M. C., Rouzina, I. & Bloomfield, V. A. Thermodynamics of DNA interactions from single molecule stretching experiments. *Acc. Chem. Res.* **35**, 159–166 (2002).
- Soong, R. K. *et al.* Powering an inorganic nanodevice with a biomolecular motor. *Science* **290**, 1555–1558 (2000).
- Yasuda, R., Noji, H., Kinoshita, K. Jr & Yoshida, M. F1-ATPase is a highly efficient molecular motor that rotates with discrete 120 degree steps. *Cell* **93**, 1117–1124 (1998).
- Seeman, N. C. DNA in a material world. *Nature* **421**, 427–431 (2003).
- Harada, Y. *et al.* Direct observation of DNA rotation during transcription by *Escherichia coli* RNA polymerase. *Nature* **409**, 113–115 (2001).
- Wobbe, C. R., Dean, F., Weissbach, L. & Hurwitz, J. *In vitro* replication of duplex circular DNA containing the simian virus 40 DNA origin site. *Proc. Natl Acad. Sci. USA* **82**, 5710–5714 (1985).
- Davenport, R. J., Wuite, G. J., Landick, R. & Bustamante, C. Single-molecule study of transcriptional pausing and arrest by *E. coli* RNA polymerase. *Science* **287**, 2497–2500 (2000).
- Davis, M. H. The slow translation and rotation of two unequal spheres in a viscous fluid. *Chem. Eng. Sci.* **24**, 1769–1776 (1969).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank E. Watson and Y. Inclán for technical assistance, E. Nogales for microscope time, and A. Vologodskii, V. Croquette, D. Bensimon, D. Collin, N. Pokala and Y. Chemla for critical readings of the manuscript and/or discussions. Z.B. is an HHMI predoctoral fellow, M.D.S. is supported by a PMMB training grant, and J.G. holds a fellowship from the Hertz Foundation. This work was supported by the NIH and DOE.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.B. (carlos@alice.berkeley.edu).

Visualization of an unstable coiled coil from the scallop myosin rod

Yu Li^{†‡}, Jerry H. Brown*, Ludmilla Reshetnikova*, Antal Blazsek§, László Farkas§, László Nyitrai§ & Carolyn Cohen*

* Rosenstiel Basic Medical Sciences Research Center, and † Biophysics and Structural Biology Program, Brandeis University, Waltham, Massachusetts 02454-9110, USA

§ Department of Biochemistry, Eötvös Loránd University, Pázmány P. s. 1/c, 1117 Budapest, Hungary

α -Helical coiled coils in muscle exemplify simplicity and economy of protein design: small variations in sequence lead to remarkable diversity in cellular functions^{1,2}. Myosin II is the key protein in muscle contraction, and the molecule's two-chain α -helical coiled-coil rod region—towards the carboxy terminus of the heavy chain—has unusual structural and dynamic features. The amino-terminal subfragment-2 (S2) domains of the rods can swing out from the thick filament backbone at a hinge in the coiled coil, allowing the two myosin 'heads' and their motor domains to interact with actin and generate tension³. Most of the S2 rod appears to be a flexible coiled coil, but studies suggest that the structure at the N-terminal region is unstable^{4–6}, and unwinding or bending of the α -helices near the head-rod junction seems necessary for many of myosin's functional properties^{7,8}. Here we show the physical basis of a

‡ Present address: Boston Biomedical Research Institute, 64 Grove Street, Watertown, Massachusetts 02472, USA.

particularly weak coiled-coil segment by determining the 2.5-Å-resolution crystal structure of a leucine-zipper-stabilized fragment of the scallop striated-muscle myosin rod adjacent to the head-rod junction. The N-terminal 14 residues are poorly ordered; the rest of the S2 segment forms a flexible coiled coil with poorly packed core residues. The unusual absence of interhelical salt bridges here exposes apolar core atoms to solvent.

The chimaeric peptide that we have crystallized (Fig. 1a) consists of Gly-Ser-His-Met at the N terminus, followed by residues 835–885 of the myosin heavy chain from the bay scallop *Argopecten irradians* (that is, the first 51 residues of the S2 domain, which we call 'S2N51'), a Gly-Ser linker, and residues 250–281 (the 'leucine zipper') of the yeast transcription factor GCN4 (see Methods). There are two independent 'molecular dimers' in the crystallographic asymmetric unit, each forming for the most part a parallel two-stranded α -helical coiled coil that is continuous through the Gly-Ser linker between S2N51 and GCN4. The two dimers follow somewhat different paths and seem to be flexible throughout their lengths (Fig. 1b), but their basic architectures are very similar. In each case there is increasing disorder from the C terminus to the N terminus. Along most of the relatively well-ordered C-terminal 63 residues of each chimaeric dimer, the interhelical distance is between 9 and 10 Å, typical of canonical two-stranded coiled coils. Closer to the N terminus (and further from the leucine zipper), from residues 857 to 849—where bulky methionine and glutamine residues are located in the core of the molecule—the crystallographic temperature factors rise and the diameter of the

coiled coil increases to 11 Å. Another pair of poorly ordered helical turns occurs between residues 849 and 846 of each chain, and the N-terminal 8–11 residues of S2, as well as the N-terminal tetrapeptide of this construct, cannot be seen in the electron-density maps.

The structure of the core of scallop S2N51 reveals several features that seem to contribute to a flexible coiled coil (Figs 2 and 3). The key signature of α -helical coiled coils is a short-range seven-residue ('heptad') sequence repeat in the form 'abcdefg', where the core 'a' and 'd' residues are generally apolar. These residues form a left-handed apolar stripe along the surface of the right-handed α -helices, and their side chains interlock in a 'knobs-into-holes' fashion by the winding of helices around one another to produce the coiled-coil structure^{1,2,9,10}. Scallop S2N51, despite the uninterrupted heptad repeat, has a large proportion of polar or charged residues in the core, especially at the N terminus. Here, the first three nominal 'd' positions of the sequence are occupied by a proline and two glutamine residues. Moreover, in each dimer of the asymmetric unit, the two 'd'-position Gln 849 residues pack in the core in an asymmetric fashion, with only one of the side chains forming a knob-into-hole contact with the opposite helix; the other side chain is oriented towards the solvent. (Asymmetrically packed glutamine residues also mark a boundary between coiled-coil and non-coiled-coil regions of tropomyosin¹¹.) These sequence features

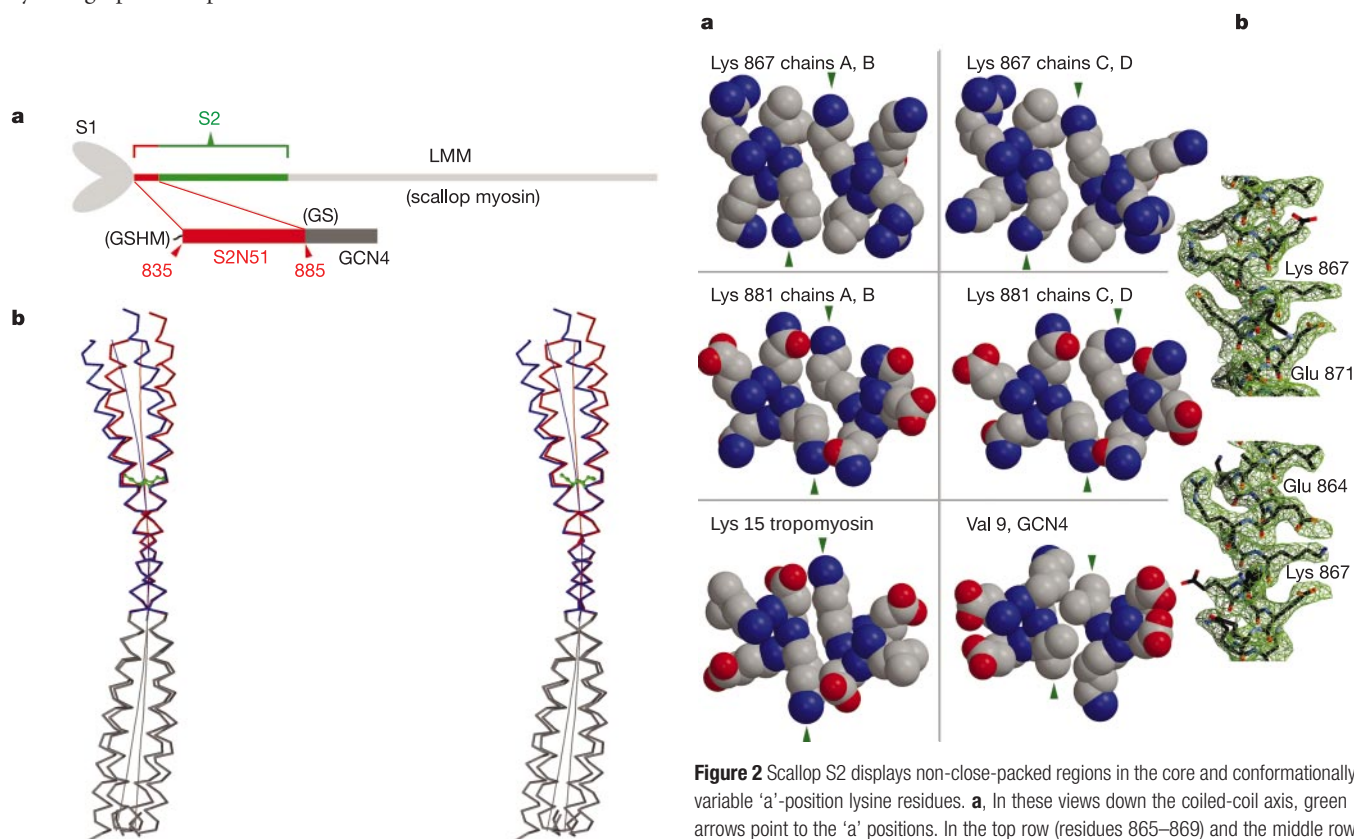


Figure 1 S2N51 is a flexible coiled coil. **a**, Diagrams of a myosin II dimer (top) and an expanded view of the construct crystallized in the present study (bottom). **b**, The two dimers of the crystallographic asymmetric unit are superimposed by fitting S2 residues 870–878 (S2N51 segments as red and blue C α traces, shown in stereo). The structure near residue 867 (green) is particularly variable; here the coiled-coil axis (curved line) bends by only $\sim 1^\circ$ in one molecular dimer of the asymmetric unit (red) and by $\sim 5^\circ$ in the other (blue). Elsewhere along S2N51 both coiled coils bend by at least 3° . This continuous bending contrasts with the localized 'alanine' bending observed in tropomyosin³⁰ and the relative lack of bending ($\sim 2^\circ$) along the GCN4 portion of the dimer.

Figure 2 Scallop S2 displays non-close-packed regions in the core and conformationally variable 'a'-position lysine residues. **a**, In these views down the coiled-coil axis, green arrows point to the 'a' positions. In the top row (residues 865–869) and the middle row (residues 879–882) both dimers of the crystallographic asymmetric unit are displayed. The terminus of the Lys 867 side chain is generally oriented away from the core and from the unusual 'g'-position Ile 866, and instead interacts intrahelically with the 'e'-position glutamate residues 864 or 871 (shown in **b**) (see also Fig. 1b and the text). Lys 881 and Gln 880 form a core layer that is close packed only in one dimer of the asymmetric unit, where the lysine C γ and C δ atoms are puckered and nestle into the hole of the opposite helix. Also shown, for contrast, are a well-packed canonical knobs-into-holes 'a' layer of GCN4, and a 'g-a' salt-bridge-stabilized layer of tropomyosin (bottom row). **b**, Region near Lys 867 in helix A (top) and helix D (bottom) viewed perpendicular to the axis. The $2|F_o| - |F_c|$ electron-density map is contoured at 1σ .

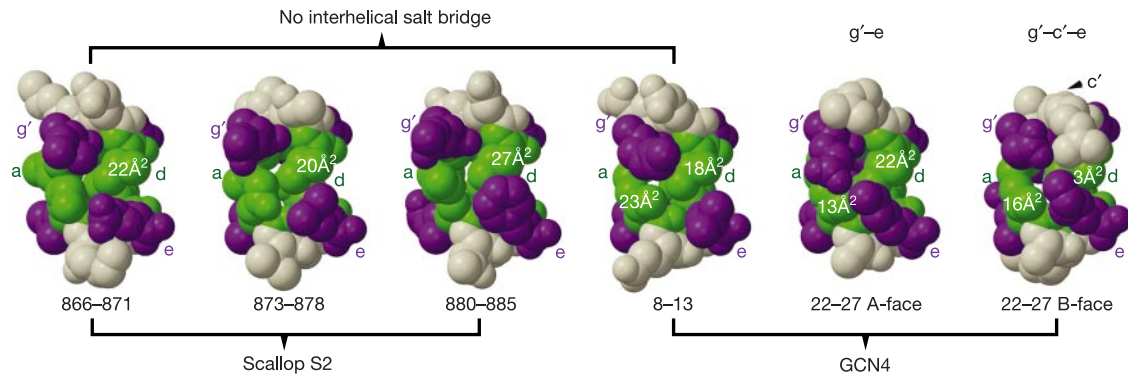


Figure 3 The absence of interhelical salt bridges between the 'g' and 'e' positions (purple) exposes core residues (green) to solvent. Each panel shows a horizontally oriented fragment of the coiled coil spanning heptad positions 'g', 'a', 'b', 'c', 'd' and 'e' from left to right. Solvent-accessible areas (in Å²) calculated by SURFACE³¹ are indicated for the apolar core residues (six 'd'-position leucines and three 'a'-position valines); other apolar

core residues of S2N51 (not shown) are also highly exposed. Also note that depending on side-chain conformations, simple 'g-e' salt bridges may bury either the intervening 'a' position (for example in the fifth panel) or the 'd'-position residue (for example leucine B39 of tropomyosin³⁰, solvent-accessible area 3 Å²).

are consistent with the poor order in the density maps at the N-terminal 14 residues of scallop S2. In the remainder of scallop S2N51, 4 of the 10 core positions are polar (Fig. 4). The unbranched long lysine side chains in the 'a' positions 867 and 881 also adopt variable conformations, which form different sets of interhelical and intrahelical interactions on the surface of the coiled coil (see Fig. 2). Lys 867 is in fact adjacent to Ile 866, whose apolar and β-branched side chain is unusual for a 'g' position; these residues together form a poorly packed core for all four helices of the asymmetric unit (Fig. 2). Moreover, at this location the two molecules of the asymmetric unit follow distinctly different paths (Fig. 1b). (Note that a mutation from arginine to histidine at the corresponding position (870) of cardiac β-myosin is associated with familial hypertrophic cardiomyopathy¹²; because residue 867 is exposed in our structure and

histidine is a less flexible side chain than arginine or lysine, we expect that this mutation might alter both the dynamic and interaction properties of the molecule.) The only canonical region in the core of S2N51 is near its C terminus, which contains three consecutive relatively conserved 'd'-position leucines (Fig. 4a) (but see below). As is found in other coiled-coil proteins, the canonical small or branched apolar core side chains can adopt only a few (generally two) rotamers and they consistently form interhelical knob-into-hole hydrophobic contacts, in contrast to the conformationally variable 'a'-position lysines.

Scallop S2N51 is also remarkable in lacking any interhelical salt bridges. Coiled coils seem to require intrahelical and interhelical salt bridges for their stability¹³⁻¹⁵, especially those between the flanking 'e' and 'g' positions of neighbouring chains^{16,17}. The first potential

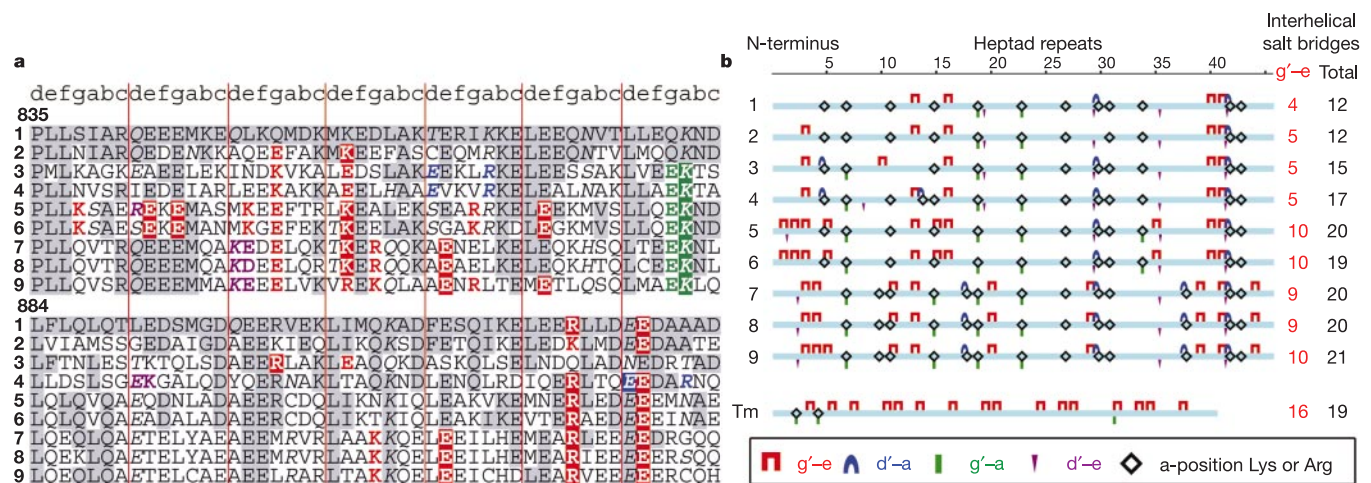


Figure 4 The N-terminal region of molluscan S2 (sequences 1 and 2) displays in an exaggerated way the moderately low level of coiled-coil-stabilizing features present in S2 of various myosins. The sequences of the N-terminal 14 heptads (a) and schematic summaries of S2 domains (b) from invertebrate (sequences 1-4), vertebrate striated muscle (sequences 5 and 6), vertebrate smooth muscle (sequences 7 and 8) and vertebrate non-muscle myosins II (sequence 9). The S2 sequences shown are from the striated adductor muscle of the bay scallop *Argopecten irradians* (GenBank accession no. 5612) (1), the ventral siphon of the squid *Loligo pealei* (GenBank 3252880) (2), the body wall muscle of *Caenorhabditis elegans* (unc-54, GenBank 17509401) (3), indirect flight muscle of *Drosophila melanogaster* (GenBank 20455497) (4), human cardiac muscle

(β-myosin, GenBank 547966) (5), chicken (*Gallus gallus*) skeletal muscle (GenBank 6683485) (6), chicken-gizzard muscle (GenBank 86369) (7), human bladder and oesophagus muscles (GenBank 13432177) (8), and human platelet (GenBank 6166599) (9). For comparison, a schematic of the coiled coil of rat (*Rattus norvegicus*) striated-muscle α-tropomyosin (Tm) is also shown. In a, identities with the scallop sequence are shaded, core polar side chains (generally coiled-coil destabilizing) are italicized; in b, 'a'-position lysines and arginines are indicated by black diamonds, and in both panels the residues participating in interhelical 'g-a', 'd-e', 'd-a' and the well-recognized^{16,17} coiled-coil stabilizing 'g-e' salt bridges are coloured green, magenta, blue and red, respectively.

interhelical salt bridge in scallop S2 occurs 88 residues from the head-rod junction, and the S2 regions of molluscan myosins have fewer potential interhelical salt bridges than do those of other conventional myosins (Fig. 4). (Note that this segment does have intrahelical salt bridges in numbers typical of other two-stranded coiled coils.) A comparison of the S2N51 structure with that of other coiled coils (Fig. 3) illustrates how 'g-e' (and especially complex 'g-c-e') salt bridges, in addition to providing ionic bonding between the two α -helices, can also contribute to the stability of a coiled coil by helping to bury residues in the core (see also ref. 17). The residues that are exposed in scallop S2N51 because of the lack of 'g-e' links include the three consecutive 'd'-position leucines (Figs 3 and 4) as well as other apolar core side chains. Removal of apolar residues from aqueous solvent is a key driving force of protein folding, so that the stability of coiled coils might not simply be a function of the number of good (that is, apolar, leucine) 'a' and 'd' residues and the number of interhelical ('g-e') salt bridges, but also of their specific locations relative to one another. The likely importance of the sequence context of charge pairs has been indicated by recent mutational studies of heterodimeric coiled coils¹⁸. A definite positional relationship in the primary sequence between locally clustered salt bridges and hydrophobic core residues is also a key feature of 'trigger sequences', which promote α -helical coiled-coil formation in several proteins¹⁴.

The structure of the N-terminal 51 residues of S2 in scallop displays in an exaggerated way properties of the entire S2 domain. The sequences of S2 and the rest of the rod in diverse isoforms generally display somewhat fewer potential interhelical salt bridges (~0.3 per heptad)¹⁹ and apolar residues in the core (~65%) as well as significantly more 'a'-position lysines and arginines (~20–25%)²⁰ than do most α -helical dimeric coiled coils^{19–21} (see Fig. 4). Coiled-coil trigger sequences containing multiple potential salt bridges are also generally absent from the N-terminal region of S2. (In scallop S2 there is a trigger sequence¹⁴ located 85 residues from the head-rod junction, but it is embedded in a region of poor coiled-coil potential.) Correspondingly, in contrast to particularly stable coiled coils with triggers such as those in corticillin I (ref. 15) and the 33-residue GCN4 leucine zipper¹⁰, relatively long N-terminal segments of S2 are often required for dimerization (more than 300 residues in scallop⁶). The extent of S2 instability seems to vary depending on the isoform^{4–6,12}.

Weakness in the N-terminal region of the S2 coiled coil seems to be an important aspect of its design. The disorder we have observed in the first heptad of the scallop structure is characteristic, in fact, of all conventional myosin isoforms, and is signalled by the conserved proline present at the head-rod junction. This structural feature accounts for the pivot needed for mobility of the head-rod junction. In all muscles, the S2 region must be strong enough to transmit tension efficiently during contraction, but localized weakness could contribute to the compliance of the myosin crossbridge. Mutational studies of chicken-gizzard myosin heavy meromyosin (HMM) indicate that instability at the head-rod junction might be important for optimal mechanical performance⁷. A locally weak dimeric S2 coiled coil might also be necessary to produce the 'off state' of regulated myosins. In myosins from vertebrate smooth muscles, regulation is controlled by the phosphorylation of a light chain in each of the heads of the molecule, whereas in myosins from scallop and other molluscan striated muscles control is effected by the direct binding of Ca^{2+} to the essential light chain²². The unphosphorylated off state of chicken smooth-muscle HMM and myosin has been visualized using cryoelectron microscopy of two-dimensional crystals: the molecules have a compact conformation with asymmetric interactions between the two heads⁸. By negative staining, the rod region has been shown to extend back from the S1–S2 junction and to run between the two heads²³. This arrangement would appear to require an unwinding of perhaps one to two heptads of the coiled coil at the N terminus of S2. In scallop HMM,

a similar compact conformation probably also occurs in the Ca^{2+} -free off states²⁴. Note also that in scallop myosin, where the N-terminal region of S2 seems to be especially unstable⁶, the time required for the 'switching on' of the heads is considerably shorter (within 10 ms (ref. 25)) than that for vertebrate smooth muscles. In contrast to these results on regulated myosins, it seems that the extent of instability in S2 is less in vertebrate skeletal-muscle myosin, because resonance-energy-transfer experiments indicate that in this S2, only the first heptad may be uncoiled when the two heads are bound to actin in the rigor state²⁶. (In this regard, note the additional potential interhelical salt bridges near the N terminus of vertebrate skeletal S2 compared with regulated myosins (Fig. 4).) These various results use different methods, isoforms and states of contraction, but are beginning to yield crucial information to relate sequence to function.

A poorly stabilized coiled coil adjacent to the motor domains in the heads is a design that might also be used in kinesin (but see ref. 27) and in various unconventional myosins. Here, coiled-coil unwinding might help to account for otherwise puzzling properties of myosin VI, which—in contrast to conventional myosin II—has a short single IQ-repeat lever arm but is a processive two-headed motor with a large (possibly 36-nm) step size²⁸. The N-terminal 60 residues of this coiled coil are also predicted to lack interhelical salt bridges. Taken together, these results bear out the versatility of the simple coiled-coil motif, which has both structural and dynamic roles in protein function²⁹. □

Methods

For details of expression, purification, crystallization and structure determination, see Supplementary Methods. In brief, the chimaeric peptide (see Fig. 1a) was expressed in *Escherichia coli*, purified sequentially on nickel-affinity, MonoQ ion-exchange and Superdex 200 columns (Pharmacia), and was crystallized after repeated seedings at 16 °C against 15–20% poly(ethylene glycol) 2000 monomethyl ether, 50 mM NaCl, 2 mM Na₂S₂O₈, 44 mM MOPS pH 6.9. X-ray data from a single crystal (cryopreserved with glycerol) were collected at 100 K ($P2_12_12_1$, $a = 54.04$ Å, $b = 73.30$ Å, $c = 102.97$ Å, 49% solvent). The structure was determined by a novel use of molecular replacement in which some conventional refinement was performed on a correctly oriented model before the calculation of the translation function (final $R_{\text{free}} = 0.286$ to 2.5 Å resolution; root-mean-square bond lengths and angles are 0.006 Å and 1.012°, respectively). Note that the GCN4 leucine zipper portion of the chimaeric peptide promotes the formation of the S2N51 coiled coil (otherwise transiently stable⁹) but is not expected to perturb its structure owing to the linear nature of the coiled-coil motif.

Received 6 January; accepted 20 May 2003; doi:10.1038/nature01801.

- Cohen, C. & Parry, D. A. D. α -Helical coiled coils and bundles: How to design an α -helical protein. *Proteins* **7**, 1–15 (1990).
- Lupas, A. Coiled coils: New structures and new functions. *Trends Biochem. Sci.* **21**, 375–382 (1996).
- Huxley, H. E. The mechanism of muscular contraction. *Science* **164**, 1356–1365 (1969).
- Trybus, K. M. Regulation of expressed truncated smooth muscle myosins. Role of the essential light chain and tail length. *J. Biol. Chem.* **269**, 20819–20822 (1994).
- Knight, P. J. Dynamic behavior of the head-tail junction of myosin. *J. Mol. Biol.* **255**, 269–274 (1996).
- Mahnasi-Csizmadia, A., Shimony, E., Hegyi, G., Szent-Györgyi, A. G. & Nyitrai, L. Dimerization of the head-rod junction of scallop myosin. *Biochem. Biophys. Res. Commun.* **252**, 595–601 (1998).
- Lauzon, A. M., Fagnant, P. M., Warshaw, D. M. & Trybus, K. M. Coiled-coil unwinding at the smooth muscle myosin head-rod junction is required for optimal mechanical performance. *Biophys. J.* **80**, 1900–1904 (2001).
- Liu, J., Wendt, T., Taylor, D. & Taylor, K. Refined model of the 10S conformation of smooth muscle myosin by cryo-electron microscopy 3D image reconstruction. *J. Mol. Biol.* **329**, 963–972 (2003).
- Crick, F. H. C. The packing of α helices: simple coiled-coils. *Acta Crystallogr.* **6**, 689–697 (1953).
- O'Shea, E. K., Klemm, J. D., Kim, P. S. & Alber, T. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* **254**, 539–544 (1991).
- Li, Y. *et al.* The crystal structure of the C-terminal fragment of striated-muscle α -tropomyosin reveals a key troponin T recognition site. *Proc. Natl Acad. Sci. USA* **99**, 7378–7383 (2002).
- Gruen, M. & Gautel, M. Mutations in β -myosin S2 that cause familial hypertrophic cardiomyopathy (FHC) abolish the interaction with the regulatory domain of myosin-binding protein-C. *J. Mol. Biol.* **286**, 933–949 (1999).
- McLachlan, A. D. & Stewart, M. Tropomyosin coiled-coil interactions: Evidence for an unstaggered structure. *J. Mol. Biol.* **98**, 293–304 (1975).
- Kammerer, R. A. *et al.* An autonomous folding unit mediates the assembly of two-stranded coiled coils. *Proc. Natl Acad. Sci. USA* **95**, 13419–13424 (1998).
- Burkhardt, P., Kammerer, R. A., Steinmetz, M. O., Bourenkov, G. P. & Aebi, U. The coiled-coil trigger site of the rod domain of corticillin I unveils a distinct network of interhelical and intrahelical salt bridges. *Structure* **8**, 223–230 (2000).
- Meier, M., Lustig, A., Aebi, U. & Burkhardt, P. Removing an interhelical salt bridge abolishes coiled-coil formation in a *de novo* designed peptide. *J. Struct. Biol.* **137**, 65–72 (2002).
- Burkhardt, P., Ivaninskii, S. & Lustig, A. Improving coiled-coil stability by optimizing ionic interactions. *J. Mol. Biol.* **318**, 901–910 (2002).

18. Arndt, K., Pelletier, J., Muller, K., Pluckthun, A. & Alber, T. Comparison of *in vivo* selection and rational design of heterodimeric coiled coils. *Structure* **10**, 1235–1248 (2002).
19. Conway, J. F. & Parry, D. A. D. Structural features in the heptad substructure and longer range repeats of two-stranded α -fibrous proteins. *Int. J. Biol. Macromol.* **12**, 328–334 (1990).
20. McLachlan, A. D. & Karn, J. Periodic features in the amino acid sequence of nematode myosin rod. *J. Mol. Biol.* **164**, 605–626 (1983).
21. Woolfson, D. N. & Alber, T. Predicting oligomerization states of coiled coils. *Protein Sci.* **4**, 1596–1607 (1995).
22. Adelstein, R. S. & Eisenberg, E. Regulation and kinetics of the actin–myosin–ATP interaction. *Annu. Rev. Biochem.* **49**, 921–956 (1980).
23. Burgess, S. *et al.* Structure of smooth muscle myosin in the switched-off state. *Biophys. J.* (annual meeting abstracts), 356a (2002).
24. Stafford, W. F. *et al.* Calcium-dependent structural changes in scallop heavy meromyosin. *J. Mol. Biol.* **307**, 137–147 (2001).
25. Nyitrai, M., Szent-Györgyi, A. G. & Geeves, M. A. A kinetic model of the co-operative binding of calcium and ADP to scallop (*Argopecten irradians*) heavy meromyosin. *Biochem. J.* **365**, 19–30 (2002).
26. Chakrabarty, T., Xiao, M., Cooke, R. & Selvin, P. R. Holding two heads together: Stability of the myosin II rod measured by resonance energy transfer between the heads. *Proc. Natl Acad. Sci. USA* **99**, 6011–6016 (2002).
27. Tomishige, M. & Vale, R. D. Controlling kinesin by reversible disulfide cross-linking. Identifying the motility-producing conformational change. *J. Cell Biol.* **151**, 1081–1092 (2000).
28. Buss, F., Luzio, J. P. & Kendrick-Jones, J. Myosin VI, a new force in clathrin mediated endocytosis. *FEBS Lett.* **508**, 295–299 (2001).
29. Burkhard, P., Stetefeld, J. & Strelkov, S. V. Coiled coils: A highly versatile protein folding motif. *Trends Cell Biol.* **11**, 82–88 (2001).
30. Brown, J. H. *et al.* Deciphering the design of the tropomyosin molecule. *Proc. Natl Acad. Sci. USA* **98**, 8496–8501 (2001).
31. Collaborative Computational Project, Number 4. *Acta Crystallogr. D* **50**, 760–763 (1994).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank D. A. D. Parry, A. G. Szent-Györgyi and H. E. Huxley for a critical reading of the manuscript, and the staff of the Cornell High Energy Synchrotron Source for assistance with data collection. This work has been supported by grants to C.C. from the National Institutes of Health and the Muscular Dystrophy Association, and to L.N. from the Hungarian Scientific Research Fund (OTKA).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence should be addressed to (ccohen@brandeis.edu). X-ray coordinates have been deposited at the Protein Data Bank, Brookhaven (1NKN).

Cell, cell, cell

New for tissue and cell culture.



Eppendorf 5702R: ready for a spin.

Centrifuge 5702R

Eppendorf www.eppendorf.com
All-round coolness

The new refrigerated bench-top Centrifuge 5702 R fits in the Eppendorf model range between the 5417 R and the 5804 R. Designed for small medical or tissue cell culture labs, the 5702 R can maintain the temperature in the range -9°C to 40°C , at runs safely at maximum rotational speed at 4°C . Time-saving features include refrigeration in standby mode and a 'Fast-temp' setting. Standard settings can be saved and a number of exclusive functions widen the range of possible applications. Two program keys allow for the saving of standards settings and special functions. Three different rotors are available — one fixed-angle and two swing-bucket — designed to combine useful capacity and simple handling. All rotors and accessories are autoclavable (121°C , 20 min). The maintenance-free rotor allows for a short acceleration time of less than 25 seconds to maximum rotor speed.

FreeStyle 293 Expression System

Invitrogen www.invitrogen.com
A good source of protein

Invitrogen's new FreeStyle 293 Expression System transfects 293 cells in a single suspension culture, producing greater protein yields than generally obtained using adherent cell culture techniques. This system eliminates the need to feed, pass and maintain multiple flasks. The media used, which

need no changing, are serum-free, so the problems associated with downstream purification are reduced.

Chick embryo extract

Accurate Chemical & Scientific

www.accuratechemical.com

CEE double

Chick embryo extract (CEE) from Accurate Chemical is now available in two formulations — a frozen solution and as a lyophilized powder. The extract, which is suitable for *in vitro* cell culture, is processed from eggs collected from registered flocks that are free from clinical signs of notifiable diseases.

Process Pro-280

Ismatec/Micropump

www.ismatec.com/ www.micropump.com

Narrow escape

The Process Pro-280 peristaltic pump system consists of a drive and a pump-head and features a stainless steel convex roller and concave tub-bed design that promotes gentle pumping. Before the roller totally closes the squeezed tubing, the cells can escape through a temporary gap created towards the tubing wall and thus are not damaged. The manufacturer points to studies showing that cell viability as high as 87% after 56 hours of continuous transfer can be achieved using the Pro-280. The Process drive comes with the IP65 protection rating suitable for clean room areas. Other features include interchangeability with more than 20 available single- and multi-channel pump heads offering flow rates from 0.001 to 3700 ml/min, a self-centring tube-track and an RS232/analogue interface. It also has four program memories and is programmable for stand-alone process operation. Applications include pumping highly concentrated cell cultures in biotechnology and pharmaceutical research, including laboratory scale and small production scale-up.

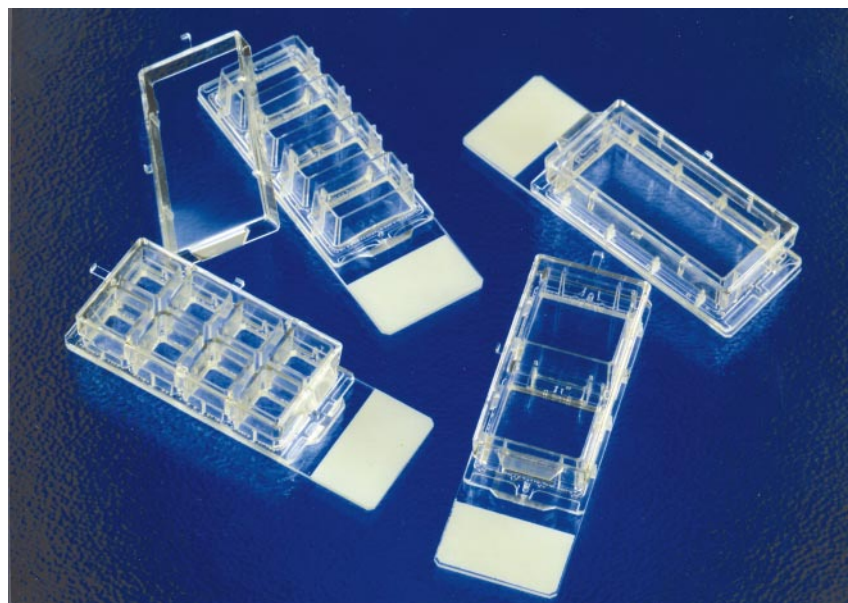
Iwaki slides

Bibby Sterilin

www.bibby-sterilin.com

For experiments conducted in chambers

Iwaki chamber slides from Bibby Sterilin enable the user to culture, fix, stain and observe cells, all on one slide. They are available in chamber formats of one, two, four and eight. Each individual chamber allows simultaneous multicultures. Viewing of the cells is said to be easier and more accurate than with conventional culture plates and flasks. A non-toxic silicone adhesive attaches



Chamber performance: Iwaki slides, for small-scale cultures.

the polystyrene chambers to the glass base, and for viewing under a microscope they are simply manually removed from the glass slide. The chamber slides are especially suitable for promoting uniform growth of small cultures for research purposes and in cases where only a minute sample is available.

Scientific Reports online

Integra Biosciences

www.integra-biosciences.com

On-line cell cultivation bibliography

This new on-line service from Integra Biosciences lists over 100 reference papers from around the world covering research, laboratory-scale and full production-scale applications performed using membrane and hollow-fibre technologies for monoclonal antibody production or cell cultivation. Researchers can use it to obtain references for their work and can add their own papers.

The bibliography is organized by cell and antibody type, and can be found at www.integra-biosciences.com/e-fhomecc.html

CD hybridoma and CHO media

Invitrogen

www.invitrogen.com

Dry cell culture media in granular form

Invitrogen's GIBCO chemically-defined (CD) hybridoma and Chinese hamster ovary (CHO) media have recently become available in the company's Advanced Granulation Technology (AGT) format. AGT media are dry cell culture media provided in a new, easy-to-use granular format that reduces medium preparation time and can be used for many complex formulations. CD hybridoma medium is optimized for hybridoma growth and monoclonal antibody production. It is suitable for the culture of recombinant myeloma lines as well as



A close-up of GIBCO's granular CHO medium.

traditional hybridomas. Although designed for batch systems, it can be modified easily for use in others. CHO medium is optimized for the growth of CHO cells and expression of recombinant proteins or monoclonal antibodies in suspension culture.

CELLine

Integra Biosciences

www.integra-biosciences.com

Growth industry

The CELLine cultivation system features semi-permeable membranes above and below the cell cultivation chamber that allow nutrients and gases to diffuse. Separate ports provide selective access to the upper nutrient supply chamber and the central cultivation chamber. Medium can be exchanged without influencing the function or growth of the cells and gases. Consequently, cell growth limitations brought about by a lack of nutrients or accumulation of metabolic waste can be overcome. The system enables researchers to use higher cell concentrations in the cultivation chamber and obtain highly concentrated cell products. The manufacturer claims that, by optimizing growth conditions in this way, cell densities can be significantly increased, even to the extent of three-dimensional cell growth, and a typical cost or reduction of 40% can be achieved. The transparent design of the CELLine system provides visual control and microscopic observation of the growth process. It also offers cultivation volumes of 0.75–15 ml, making it suitable for both research and pilot plant cell cultivation processes.

Krystal clear microplates

Porvair Sciences www.porvair-sciences.com

The bottom line

With a choice of designs and formats including 24-, 96- and 384-wells, a Krystal clear-bottomed microplate is now available to suit almost all photometric and cell counting applications. The 384-well Krystal plate

range features what the makers claim is market-leading plate flatness of 50.10 mm. For scientists performing cell based assays with fluorescent imagers this translates into elimination of misreads and a significant increase in measurement precision. The clear cup design of Krystal 2000 96-well plates ensures zero well-to-well light cross talk, and high cell binding efficiency together with the convenience of direct use on inverted microscopes and bottom reading spectrophotometers. Available in standard white or black, high bind or tissue culture treated versions Krystal 96 plates offer a versatile 96-well solution for most photometric applications. Krystal range microplates are fully compatible with all automated liquid handling systems, photometric readers and robotic handling devices.

Cryogenic vials

Corning

www.corning.com

Better threads on the block

Corning's range of internal and external thread cryogenic vials—1.2 ml, 2.0 ml, 4.0 ml and 5.0 ml—is now certified RNase- and DNase-free. In addition, the vials are guaranteed nonpyrogenic, and are supplied sterile (sterilization is by gamma irradiation).



Vial bodies: Corning's range, nuclease free.

They have quicker volume identification, featuring a 40% larger marking area to make it easier to read labels. Other convenient improvements include easier and more secure storage, sample tracking and better identification. These cryogenic vials can still withstand temperatures as low as -196°C , and self-standing vials are still available for safe, single-handed use. Internal and external thread vials come with totally secure sealing. A well-improved and useful product for maximum efficiency.

These notes are compiled in the Nature office from information provided by the manufacturers.

Contacts

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Global Head of Naturejobs:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)
Andy Douglas (4975)
Frank Phelan (4944)

Netherlands/ Italy/ Iberia/ Belgium:

Evelina Rubio Hakansson (4973)
Scandinavia: Silje Opstrup (4994)

France/ Switzerland:
Amelie Pequignot (4974)

Natureevents:
Paul Constant (4954)

Production Manager: Billie Franklin

To send materials use London
address above.
Tel +44 (0) 20 7843 4814
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

International

Advertising Coordinator:
Hind Berrada (4935)

Naturejobs web development:
Tom Hancock

Naturejobs online production:
Ben Lund

European Satellite Office

Germany/ Austria:

Patrick Phelan, Odo Wulffen
Tel + 49 89 54 90 57 11/2
Fax + 49 89 54 90 57 20
e-mail: p.phelan@nature.com
o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

US Advertising Coordinator:
Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Harakutamachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746

Asia-Pacific Sales Director:
Rinoko Asami
e-mail: r.asami@naturejp.com

naturejobs

Golden handcuffs

Mexico used a man in an over-sized sombrero, Manitoba had a stuffed bear, and Wisconsin deployed a Harley-Davidson motorcycle. At the Biotechnology Industry Organization's annual conference in Washington DC last month, countries, provinces and states went to great lengths to attract people to their stands.

Singapore was no exception. The country sponsored a large hut-like booth, and used free food to entice passers-by, whose goodie bags already overflowed with free pencils, foam toys and slinkies. But the country is going to even greater lengths to stop its most promising talent from leaving. Every year it selects 90 of its best students and funds them — complete with travel allowance and generous monthly stipend — from their undergraduate days through to their postdocs.

The students can study wherever they please — and they are encouraged to go to the world's best institutions. But when they have finished training, they must spend six years in Singapore at a national research institute or government-sponsored lab.

Although this model may not work in countries with less centralized governments, it has some appealing aspects. The idea of recognizing promise early and committing to it is commendable. Other models, such as loan forgiveness for young scientists who stay in their home countries, might serve the same purpose. Another possibility is the provision of start-up research packages for young investigators who don't yet have enough data to publish their breakthrough paper, but who show signs of being able to do so, given the right infrastructure and support.

Singapore's presence in the biotech exhibition hall sends a message: toys are great for attracting attention, but commitment to facilities and funding is the best way to recruit young scientists.

Paul Smaglik

Naturejobs editor



Contents

REGIONS

A change of pace in
western Japan

p352

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

Changing gear: Japanese research is seeking to become more dynamic, flexible and industry-friendly.



Go west, young postdoc

Western Japan

Japan's political and economic power is largely concentrated in the greater Tokyo region on Honshu Island's east coast. But much of its research strength, especially in biological and medical sciences, lies in western Japan, stretching from Osaka and Kyoto to the southern island of Kyushu. Many of the region's immunologists and developmental biologists, for example, are world leaders, and five of Japan's nine natural-science Nobel prizes were won by Kyoto University graduates.

But leaders in the region have not been content to live on bygone glories. The past five years have been marked by generous initiatives to build and invigorate academic research. The changes, in combination with university reforms, are meant to make Japanese academic research younger, more dynamic, flexible and industry-friendly. Research leaders are looking to foreign scientists and industry to help them to achieve those ends.

Institutions in western Japan are spearheading these initiatives. A research community is taking shape in the region, where young scientists can find well-funded, independent positions that do not exist elsewhere — although this is an ongoing process. Two years ago, Kyoto University's medical school received a rap on the knuckles in one of the education ministry's first

external reviews. Although it was credited with having world-leading researchers, it was criticized for having few independent leaders under 40 years old. Since then, a grant from the education ministry has helped the school add 22 group leaders, with an average age of 35, to a staff of 66 principal investigators as part of a five-year trial project that aims to switch the emphasis away from seniority and hierarchy.

YOUNG AND FREE

It has been a challenge to integrate these new faces with the old research structure because the traditions of the seniority system are taking a long time to disappear. But slowly the new and old researchers have found a way to help each other, says immunologist Tasuku Honjo, dean of Kyoto's medical school. If, after five years, these recruits have proved themselves and find other positions, at the university or elsewhere, more opportunities for young researchers are likely to follow. "It's a risk," says Honjo. "But if Kyoto University changes, others will likely change too."

Regional universities in the Kyoto–Osaka area are also pioneering efforts to break down the rigidity of departmental barriers, to allow students to take

advantage of all the talent in the university rather than working with a single professor. Following the example set by Kyoto University's Graduate School of Biostudies, Osaka University's Graduate School for Frontier Biosciences, in its second year, will give graduate students a chance to work with high-profile biologists before deciding on their graduate course.

These established universities have had to change to compete with newer institutions. The well-funded RIKEN Center for Developmental Biology in Kobe, which was established two years ago (see *Nature* **415**, 952–953; 2002), and the Nara Institute of Science and Technology (NAIST) have also set a premium on youth and independence. Although small, NAIST has pioneered a new model. Its spacious, well-equipped labs have attracted a lot of talent in its 12-year history. It recently claimed two of the country's prestigious — if financially modest — centre of excellence (COE) awards, which provide between US\$1 million and US\$2 million per year for five years for multidisciplinary projects.

The COE programme has added impetus to Japan's efforts to create a class of postdocs — a position that is relatively new to Japan. NAIST's life-science COE added 20 postdoc students, doubling the previous number. Some 100 additional COE recipients will be selected this month.

The COE scheme offers a chance for smaller universities to gather talent, and for young researchers to find interesting short-term projects outside the former imperial universities, says Miyazaki Medical College president Hisayuki Matsuo. "Even small universities with well-organized groups can make a mark in certain fields," he says. Shizuo Akira, an immunologist at Osaka University, agrees: "There was a time when everyone who could, automatically went to Tokyo University. Now students are beginning to evaluate which universities are strong for which fields."

The universities are aided in their efforts to recruit



Osaka University is encouraging students to work with world leaders in their fields.

young or overseas researchers by government funding agencies. The Japan Science and Technology Corporation, for example, sponsors major projects that support many postdoctoral students. And the Japan Society for the Promotion of Science awards grants specifically to foreigners.

FOREIGNER-FRIENDLY?

But do foreigners want to come to Japan? Funding is generous, and changes in the research structure are making it easier for them to do what they want. Japan's culture is becoming more open to foreigners, and the pace of life in the west is gentle. If there is a conference in Tokyo, the capital is just a few hours away by bullet train.

Yet very few foreigners are taking up these positions. One concern is that after their three- or five-year contract runs out, they will be left without a place to go. This is slowly changing as research initiatives and universities turn towards short-term contracts rather than lifelong jobs. In time, it should become easier for foreigners to integrate into the system.

Some women are wary of attempting to advance their careers in Japan. But although barriers that have hampered Japanese women might be slow to collapse, foreign women generally do not have a problem.

Communication is not a problem with most senior researchers, but it can be quite difficult to discuss scientific experiments with younger colleagues, says Damien Lieberherr, a Swiss postdoctoral fellow at NAIST. But for Romanian immunologist Sidonia Fagarasan, the facilities at her disposal at Kyoto University's graduate medical school were a "dream" that eclipsed any language difficulties, she says.

More efforts to make Japan more foreigner-friendly are under way, with plans for one of the largest projects beginning in Osaka. A new institute encompassing life sciences, information technology and nanotechnology plans to have foreigners make up half of its staff when it opens in 2007.

Osaka University president Tadamitsu Kishimoto admits it is a problem to recruit foreigners when "every document is in Japanese". But he hopes this will change with the nationwide university reforms next year, which will even allow foreigners to become university presidents. "We need some foreigners in top positions — like Nissan or the Japanese soccer team did — to shake things up," Kishimoto argues.

With the youth movement already under way, it seems that it will be only a matter of time before a foreign infusion of research talent follows.

David Cyranoski is *Nature's* Asia-Pacific correspondent.

Earthquake recovery

Although a 1995 earthquake left Kobe reeling (see below), its biology heritage is now being restored. The Foundation for Biomedical Research intends to establish an industrial park, based around RIKEN's Center for Developmental Biology (CDB) and the Institute of Biomedical Research and Innovation (IBRI). New buildings are popping up to house the IBRI's positron-emission tomography imaging devices and cell-processing centre. A ¥4.6-billion (US\$38.9-million) biomedical accelerator is planned, together with a radioisotope facility.

Some 34 firms have been established, from small start-ups to international companies. Australia's Stem Cell Sciences set up a branch a year ago, for example, and has collaborated with the CDB. It is developing products including drug screening assays. D.C.

